On Large Deviation Efficiency in Statistical Inference

Puhalskii, A. and Spokoiny, V.

February 21, 1996

Abstract

This paper presents a general approach to statistical problems with criteria based on probabilities of large deviations. The underlying idea, which originates from similarity in the definitions of the large deviation principle and weak convergence, is to develop a large deviation analogue of asymptotic decision theory.

We consider a sequence of statistical experiments over an arbitrary parameter set and introduce for it the concept of the large deviation principle (LDP) which parallels the concept of weak convergence of experiments. Our main result, in analogy with Le Cam's minimax theorem, states that the LDP provides an asymptotic lower bound for the sequence of appropriately defined minimax risks. We show next that the bound is tight and give a method of constructing decisions whose asymptotic risk is arbitrarily close to the bound. The construction is further specified for hypotheses testing and estimation problems.

We apply the results to a number of standard statistical models: an i.i.d. sample, regression, the change-point model and others. For each model, we check the LDP; after that, considering first a hypotheses testing problem and then an estimation problem, we calculate asymptotic minimax risks and indicate corresponding decisions.

Contents

1 Introduction

The approach to statistical problems based on considering probabilities of large deviations has been in use in statistical inference since the papers by Chernoff, 1952 and Bahadur, 1960.

Chernoff, 1952 considering the problem of discriminating between two simple hypotheses showed that, if the hypotheses are fixed, the error probabilities decrease exponentially as the sample size tends to infinity; the corresponding optimal exponent is specified by what is now known as Chernoff's function.

Basu, 1956 and Bahadur, 1960 proposed a criterion for comparing statistical estimators based on the view that the quality of an estimator is characterised by the probability that the true value of the parameter is covered by the confidence interval of given width $2c$ with centre at the estimate. If the width $2c$ is held fixed as the sample size grows, then the probabilities that the true value of the parameter is not covered again are exponentially small. The estimator giving the fastest decay is called now Bahadur efficient. It has been shown later that for the model with n i.i.d. observations from distribution P_θ this optimal rate in the class of consistent estimators is specified by the Kullback-Leibler information between measures $P_{\theta-c}$ and $P_{\theta+c}$ whereas without the consistency requirement it is related to Chernoff's function.

The ideas of Chernoff and Bahadur have been developed in various directions. Ibragimov and Radavichius, 1981, Kallenberg, 1981, Ibragimov and Khasminskii, 1981, Radavichius, 1983 and Radavichius, 1991 studied the properties of maximum likelihood estimators from the point of view of Bahadur's criterion. Fu, 1982, Borovkov and Mogulskii, 1992b and Borovkov and Mogulskii, 1992a analysed the second and higher order terms of the asymptotic expansions of Bahadur risks. Kallenberg, 1983, Rao, 1963 and Wieand, 1976 considered intermediate criteria for statistical estimators when the width of the confidence interval goes to zero with certain rate. Sievers, 1978 and Rubin and Rukhin, 1983 evaluated Bahadur risks for particular statistical models.

Lately this direction in mathematical statistics has received a new impetus, mostly in papers by Korostelev, 1993, Korostelev, 1995, see also Korostelev and Spokoiny, 1995, Korostelev and Leonov, 1995 where the classical large deviation set-up is considered in minimax nonparametric framework.

Our aim here is to give a unified treatment of statistical problems using large deviation considerations. The idea is to capitalise on analogies between large deviation theory and weak convergence theory (see Lynch and Sethuraman, 1987; Vervaat, 1988; Puhalskii, 1991) and develop a large deviation analogue of asymptotic decision theory, Strasser, 1985. The approach of invoking methods of weak convergence theory to obtain results about large deviations has proved its worth in various set-ups, Puhalskii, 1991, 1993, 1994a, 1994b, 1994c, 1995a, 1995b. We show that it can successfully be applied to statistical problems too.

We begin by defining in Section 2 the notion of the large deviation principle (LDP) for a sequence of statistical experiments. It is an analogue of the notion of weak convergence of statistical experiments and means, roughly, that the distributions of likelihood processes satisfy the large deviation principle, Varadhan, 1966; Varadhan, 1984. We illustrate the general definition on a number of standard statistical models (the Gaussian shift model, the model with i.i.d. observations, the "signal + white noise" model, the regression model with Gaussian and non-Gaussian errors, with deterministic and random designs, and the change-point model). We next give a sufficient condition for the LDP to hold. This condition is analogous to the local asymptotic normality condition introduced by LeCam, 1960.

The role played by the LDP for statistical experiments is revealed by an analogue of Le Cam's minimax theorem (which states that if statistical experiments weakly converge, then the minimax risks are asymptotically bounded below by the corresponding risk for the limit model, see LeCam, 1972, LeCam, 1986, Strasser, 1985). In Section 3, we show that the situation is similar in large deviation context: if a sequence of statistical experiments obeys the LDP, there is an asymptotic lower bound for appropriately defined minimax risks. The problem of evaluating the bound is a minimax optimization problem. Further in Section 3, we study the question of the sharpness of the lower bound. We show that it is sharp under a strengthened version of the LDP. This allows us to define large deviation (LD) efficient decisions as the ones which attain the lower bound. We give a method of obtaining nearly LD efficient decisions, i.e., those whose LD asymptotic risk is arbitrarily close to the lower bound.

Sections 4 and 5 deal with applications. Section 4 specifies the results of Section 3 for the cases of hypotheses testing and estimation problems and presents explicit constructions of nearly LD efficient decisions. In Section 5, we apply the machinery to the models introduced in Section 2: we check the LDP, give conditions when the lower bounds are attained, calculate them for hypotheses testing and estimation problems and indicate nearly LD efficient decisions. An appendix contains extensions and auxiliary results.

The results of the first four sections are new. The results we obtain for the models are partly new and partly cover or extend earlier results.

2 The Large Deviation Principle for Statistical Experiments

Let $\{\mathcal{E}_n\,, n \geq 1\}$ be a sequence of statistical experiments $\mathcal{E}_n = (\Omega_n, \mathcal{F}_n; P_{n,\theta},\ \theta \in \Theta)$ over a parameter set Θ, Strasser, 1985. In this section, we give the definition of the large deviation principle for $\{\mathcal{E}_n\,, n \geq 1\}$ and study some of its properties. We start with the case of dominated experiments.

2.1 The dominated case

Assume that each experiment $\mathcal{E}_n = (\Omega_n, \mathcal{F}_n; P_{n,\theta},\ \theta \in \Theta)$ is dominated by a probability measure P_n, i.e., $P_{n,\theta} \ll P_n$ for all $\theta \in \Theta$. We will also denote this by

$\{\mathcal{E}_n, P_n\,, n \ge 1\}$. Denote

$$Z_{n,\theta} = \left(\frac{dP_{n,\theta}}{dP_n}\right)^{1/n}, \qquad \theta \in \Theta, \tag{2.1}$$

and let $Z_{n,\Theta} = (Z_{n,\theta}\,, \theta \in \Theta)$. We submit R_+^Θ with Tihonov (product) topology so that $Z_{n,\Theta}$ is a random element of R_+^Θ; $\mathcal{L}(Z_{n,\Theta}|P_n)$ denotes the distribution of $Z_{n,\Theta}$ on R_+^Θ under P_n. The large deviation principle for $\{\mathcal{E}_n, P_n\,, n \ge 1\}$ means, roughly, that the sequence of distributions $\{\mathcal{L}\left(Z_{n,\Theta}|P_n\right), \, n \ge 1\}$ satisfies the large deviation principle on R_+^Θ.

For a precise definition, we recall some basic notions of large deviation theory. We use Varadhan's original definitions of the rate function and the large deviation principle Varadhan, 1966; Varadhan, 1984. Let S be a Hausdorff topological space. We say that a function $I : S \to [0,\infty]$ is a rate function on S if the sets $I^{-1}([0,a])$ are compact in S for all $a \ge 0$; a sequence $\{Q_n, n \ge 1\}$ of probability measures on the Borel σ-field of S is said to obey the large deviation principle (LDP) with the rate function I if

$$\varliminf_{n\to\infty} \frac{1}{n} \log Q_n(G) \ge -\inf_{x\in G} I(x),$$

for all open $G \subset S$, and

$$\varlimsup_{n\to\infty} \frac{1}{n} \log Q_n(F) \le -\inf_{x\in F} I(x),$$

for all closed $F \subset S$.

We will also be saying that I is a probability rate function if $\inf_{x\in S} I(x) = 0$. Obviously, if I appears in the LDP, it is a probability rate function.

Next, we will say that the sequence $\{\mathcal{E}_n, P_n\,, n \ge 1\}$ satisfies condition (U) if

$$(U) \qquad \lim_{H\to\infty} \varlimsup_{n\to\infty} E_n^{1/n} Z_{n,\theta}^n 1(Z_{n,\theta} > H) = 0, \; \theta \in \Theta\,.$$

Here and below we use the notations

$$E_n^{1/n}\xi = (E_n\xi)^{1/n}, \; P_n^{1/n}(A) = (P_n(A))^{1/n}.$$

Definition 2.1 *We say that a sequence $\{\mathcal{E}_n, P_n\,, n \ge 1\}$ of dominated statistical experiments obeys the dominated large deviation principle (LDP) if*

1. *the sequence $\{\mathcal{L}\left(Z_{n,\Theta}|P_n\right), \, n \ge 1\}$ obeys the LDP on R_+^Θ with some (probability) rate function I,*
2. *condition (U) holds.*

The critical part of the definition is condition 1. Condition (U) plays a subordinate though essential role. If we disregard condition (U), the definition is analogous to the definition of weak convergence of dominated statistical experiments (Strasser, 1985) which states that likelihood ratios weakly converge. The role of condition (U) will become clear shortly: it provides for the compatibility of this definition with a more general one which does not depend on the choice of dominating measures and incorporates the nondominated case too. This implies, in particular, that the lower bounds we obtain in Section 3 for a sequence of so called large deviation risks do not depend on dominating measures either (see Remark 3.2 below). Note that an analogue of condition (U) in the theory of weak convergence of statistical experiments is a consequence of weak convergence of likelihood ratios and does not have to be singled out.

Now we consider a number of statistical models which, on the one hand, show that the LDP for the likelihood ratios arises quite naturally and, on the other hand, motivate and illustrate theoretical developments below. For each model we calculate the log-likelihood ratio $\Xi_{n,\theta} = \frac{1}{n}\log\frac{dP_{n,\theta}}{dP_n}$ and give some heuristics explaining the LDP condition. The rigourous verification of the LDP for the models is deferred until Section 5. At this point we mention that if the $\Xi_{n,\theta}$ are well-defined, then, by the contraction principle, Varadhan, 1966; Varadhan, 1984, the LDP for the sequence $\{\mathcal{L}(\Xi_{n,\Theta}|P_n), n \ge 1\}$, where $\Xi_{n,\Theta} = (\Xi_{n,\theta}, \theta \in \Theta)$ and $\mathcal{L}(\Xi_{n,\Theta}|P_n)$ is the law, under P_n, of $\Xi_{n,\Theta}$ on R^Θ submitted with Tihonov topology, implies the LDP for the sequence $\{\mathcal{L}(Z_{n,\Theta}|P_n), n \ge 1\}$.

Example 2.1 *Gaussian Observations*

Let us observe a sample of n i.i.d. r.v. $\mathbf{X}_n = (X_{1,n}, \dots, X_{n,n})$ which are normally distributed with $\mathcal{N}(\theta, 1)$, $\theta \in \Theta \subset R$. For this model, $\Omega_n = R^n$ and $P_{n,\theta} = (\mathcal{N}(\theta,1))^n$, $\theta \in \Theta$. We take $P_{n,0}$ as dominating measure P_n. Then the corresponding log-likelihood ratio is of the form

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{dP_{n,\theta}}{dP_n}(\mathbf{X}_n) = \frac{1}{n}\sum_{k=1}^n (\theta X_{k,n} - \frac{1}{2}\theta^2) = \theta Y_n - \frac{1}{2}\,\theta^2.$$

where

$$Y_n = \frac{1}{n}\sum_{k=1}^n X_{k,n},\ n \ge 1.$$

The sequence $\{\mathcal{L}\left(Y_n|P_n\right), n \ge 1\}$ satisfies the LDP on R with the rate function $I^N(y) = y^2/2, y \in R$ (see, e.g., Freidlin and Wentzell, 1984). This yields by the contraction principle the LDP for the log-likelihood ratios $\Xi_{n,\theta}$.

Example 2.2 *An I.I.D. Sample*

Let us observe an i.i.d. sample $\mathbf{X}_n = (X_{1,n}, \dots, X_{n,n})$ from distribution $P_\theta, \theta \in \Theta$. We do not specify the nature of the parameter set Θ. It can be a subset of a finite-dimensional space; also the unknown distribution (or its probability density function) can be taken as θ. We assume that the family $\mathcal{P}$ is dominated by probability measure P, i.e., $P_\theta \ll P$, $\theta \in \Theta$. This model is described by the dominated experiments $\mathcal{E}_n = (\Omega_n, \mathcal{F}_n; P_{n,\theta}, \theta \in \Theta)$ with $\Omega_n = R^n$, $\mathcal{F}_n = \mathcal{B}(R^n)$, $P_{n,\theta} = P_\theta^n$, $\theta \in \Theta$, $P_n = P^n$.

We have

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{dP_{n,\theta}}{dP_n}(\mathbf{X}_n) = \sum_{k=1}^{n}\frac{1}{n}\log\frac{dP_\theta}{dP}(X_{k,n}) = \int_R \log\frac{dP_\theta}{dP}(x)F_n(dx),$$

where

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1(X_{k,n}\le x), \qquad x\in R,$$

are empirical distribution functions.

Let $\mathcal{Y}$ be the space of cumulative distribution functions on R endowed with the topology of weak convergence of corresponding probability measures. By Sanov's theorem, see Sanov, 1957, Deuschel and Stroock, 1989, 3.2.17, the sequence $\{\mathcal{L}(F_n|P_n), n\ge 1\}$ satisfies the LDP on $\mathcal{Y}$ with $I^S(F) = K(F,P)$, $F\in\mathcal{Y}$, where $K(F,P)$ is the Kullback-Leibler information:

$$K(F,P) = \begin{cases} \displaystyle\int_R \frac{dF}{dP}(x)\log\frac{dF}{dP}(x)\,P(dx), & \text{if } F\ll P,\\ \infty, & \text{otherwise.}\end{cases}$$

Denote also for $\theta\in\Theta$ and $F\in\mathcal{Y}$

$$\zeta_\theta(F) = \int_R \log\frac{dP_\theta}{dP}(x)F(dx).$$

If the density functions $\frac{dP_\theta}{dP}(x)$ are bounded from above, bounded away from zero and are continuous in x for all $\theta\in\Theta$, then, since $\Xi_{n,\theta} = \zeta_\theta(F_n)$, the contraction principle yields the LDP for the sequence $\{\Xi_{n,\Theta}, n\ge 1\}$.

Example 2.3 *"Signal + White Noise"*

We observe the stochastic process $X_n = (X_n(t), t\in[0,1])$ obeying the stochastic differential equation

$$dX_n(t) = \theta(t)dt + \frac{1}{\sqrt{n}}\,dW(t),\ 0\le t\le 1,$$

where $W = (W(t), t\in[0,1])$ is a standard Wiener process and $\theta(\cdot)$ is an unknown function which we assume to be continuous and belong to some set Θ of functions on $[0,1]$.

This model is described by the statistical experiments $\mathcal{E}_n = (\Omega_n, \mathcal{F}_n; P_{n,\theta},\ \theta\in\Theta)$, where $\Omega_n = C[0,1]$, the space of continuous functions on $[0,1]$, and $P_{n,\theta}$ is the distribution of X_n on $C[0,1]$ for given θ. We take $P_n = P_{n,0}$, where $P_{n,0}$ corresponds to the zero function $\theta(\cdot)\equiv 0$. Then $P_{n,\theta}\ll P_n$ and, moreover, by Girsanov's formula, P_n-a.s.,

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{dP_{n,\theta}}{dP_n}(X_n) = \int_0^1 \theta(t)dX_n(t) - \frac{1}{2}\int_0^1 \theta^2(t)dt. \tag{2.2}$$

Let $C[0,1]$ be submitted with uniform metric and let $C_0[0,1]$ be its subset of functions $x(\cdot)$ which are absolutely continuous w.r.t. Lebesgue measure and $x(0)=0$. Then the sequence $\{\mathcal{L}(X_n|P_n), n\geq 1\}$ satisfies the LDP on $C[0,1]$ with

$$I^W(x(\cdot)) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^1 (\dot{x}(t))^2\,dt, & \text{if } x(\cdot)\in C_0[0,1] \\ \infty, & \text{otherwise,} \end{cases}$$

where $x(\cdot)\in C[0,1]$ and $\dot{x}(t)$ denotes the derivative of $x(\cdot)$ at t (see, e.g., Freidlin and Wentzell, 1984).

Denote for a function $\theta(\cdot)\in\Theta$ and $x(\cdot)\in C_0[0,1]$

$$\zeta_\theta(x) = \int_0^1 \theta(t)dx(t) - \frac{1}{2}\int_0^1 \theta^2(t)dt$$

where the integral is understood as a Lebesgue-Stiltjes integral.

Again the log-likelihood ratio can formally be represented as $\Xi_{n,\theta} = \zeta_\theta(X_n)$. Note however that the first integral in (2.2) is an Ito integral, so the latter equality actually is valid for functions $\theta(\cdot)$ of special sort (e.g., piecewise constant or differentiable). For these functions, the contraction principle again implies the LDP for $\{\Xi_{n,\Theta}, n\geq 1\}$. A general case is studied in Section 5.

Example 2.4 *Gaussian Regression*

We are considering the regression model

$$X_{k,n} = \theta(t_{k,n}) + \xi_{k,n}, \qquad t_{k,n} = \frac{k}{n}, \quad k=1,\ldots,n, \tag{2.3}$$

where the errors $\xi_{k,n}$ are i.i.d. standard normal and $\theta(\cdot)$ is an unknown function which again is assumed to be continuous.

In this model, $\Omega_n = R^n$, $\Theta\subset C[0,1]$ and $P_{n,\theta}$ is the distribution of $\mathbf{X}_n = (X_{1,n},\ldots,X_{n,n})$ for $\theta(\cdot)$. As above, we take $P_n = P_{n,0}$. Then

$$\begin{aligned} \Xi_{n,\theta} &= \frac{1}{n}\log\frac{dP_{n,\theta}}{dP_n}(X_n) \\ &= \frac{1}{n}\sum_{k=1}^n \theta(t_{k,n})X_{k,n} - \frac{1}{n}\sum_{k=1}^n \theta^2(t_{k,n}) \\ &= \int_0^1 \theta(t)\,dX_n(t) - \frac{1}{n}\sum_{k=1}^n \theta^2\left(t_{k,n}\right), \end{aligned}$$

where

$$X_n(t) = \frac{1}{n}\sum_{k=1}^{[nt]} X_{k,n}, \qquad 0\leq t\leq 1.$$

Let $\mathcal{Y}$ be the space of right continuous with left-hand limits functions on $[0,1]$ with uniform metric (for the measurability of X_n, see Billingsley, 1968, §8).

Since the $X_{k,n}$ are $\mathcal{N}(0,1)$-distributed under P_n, the sequence $\{\mathcal{L}(X_n|P_n), n \geq 1\}$ satisfies the LDP on $\mathcal{Y}$ with I^W (see, e.g., Puhalskii, 1994a).

Since the function $\theta(\cdot)$ is continuous, we have, for large n, the approximate equality

$$\frac{1}{n}\sum_{k=1}^{n}\theta^2(t_{k,n}) \approx \int_0^1 \theta^2(t)\,dt$$

and hence $\Xi_{n,\theta} \approx \zeta_\theta(X_n)$ with the same function ζ_θ as in the previous example. In the case where the ζ_θ are moreover differentiable, integration by parts shows that the $\Xi_{n,\theta}$ are continuous functions of X_n, and the LDP for $\{\Xi_{n,\Theta}, n \geq 1\}$ follows by the contraction principle. Again, the general case is deferred until Section 5.

Example 2.5 *Non-Gaussian Regression*

We consider the same regression model (2.3) but now assume that the i.i.d. errors $\xi_{k,n}$ have distribution P with positive probability density function $p(x)$ w.r.t. Lebesgue measure on the real line. The unknown regression function $\theta(\cdot)$ is assumed to be continuous, so $\Theta \subset C[0,1]$.

As above, for a regression function $\theta(\cdot)$, we denote by $P_{n,\theta}$ the distribution of $X_n = (X_{1,n}, \ldots, X_{n,n})$. We have, with $P_n = P_{n,0}$,

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{dP_{n,\theta}}{dP_n}(X_n) = \frac{1}{n}\sum_{k=1}^{n}\log\frac{p(X_{k,n}-\theta(k/n))}{p(X_{k,n})}.$$

Introducing the empirical process $F_n = F_n(x,t)$, $x \in R$, $t \in [0,1]$, by

$$F_n(x,t) = \frac{1}{n}\sum_{k=1}^{[nt]} 1(X_{k,n} \leq x),$$

we have that

$$\Xi_{n,\theta} = \int_0^1\int_R \log\frac{p(x-\theta(t))}{p(x)}F_n(dx,dt). \tag{2.4}$$

Define $\mathcal{Y}$ as the space of cumulative distribution functions $F = F(x,t)$, $x \in R$, $t \in [0,1]$, on $R \times [0,1]$ with weak topology. Let $\mathcal{Y}_0$ be the subset of $\mathcal{Y}$ of absolutely continuous w.r.t. Lebesgue measure on $R \times [0,1]$ functions $F(x,t)$ satisfying the condition $F(\infty,t) = t$ for all $t \in [0,1]$.

By Puhalskii, 1995c, the sequence $\{\mathcal{L}(F_n|P_n), n \geq 1\}$ obeys the LDP on $\mathcal{Y}$ with the rate function $I^{SK}(F)$ given by

$$I^{SK}(F) = \begin{cases} \displaystyle\int_0^1\int_R \log\frac{p_t(x)}{p(x)}\,p_t(x)\,dx\,dt, & \text{if } F \in \mathcal{Y}_0, \\ \infty, & \text{otherwise.}\end{cases}$$

Here $p_t(x)$ is the density of F so that $F(dx,dt) = p_t(x)\,dx\,dt$.

Denote for $F \in \mathcal{Y}_0$ and $\theta \in \Theta$,

$$\zeta_\theta(F) = \int_0^1 \int_R \log \frac{p(x-\theta(t))}{p(x)} F(dx, dt).$$

Then by (2.4), $\Xi_{n,\theta} = \zeta_\theta(F_n)$ and if the log's in the integrals in the definition of the ζ_θ are bounded and continuous, the LDP for $\{\Xi_{n,\Theta}, n \geq 1\}$ holds.

Example 2.6 *The Change-Point Model*

Let us observe a sample $X_n = (X_{1,n}, \ldots, X_{n,n})$ of real valued r.v., where, for some $k_n \geq 1$, the observations $X_{1,n}, \ldots, X_{k_n,n}$ are i.i.d. with distribution P_0 and the observations $X_{k_n+1,n}, \ldots, X_{n,n}$ are i.i.d. with distribution P_1. We are assuming that P_0 and P_1 are known and k_n is unknown. Assume also that $k_n = [n\theta]$, where $\theta \in \Theta = [0,1]$. Here $\Omega_n = R^n$, and $P_{n,\theta}$ denotes the distribution of X_n for given θ.

Let probability measure P dominate P_0 and P_1, and let

$$p_0(x) = \frac{dP_0}{dP}(x),\ p_1(x) = \frac{dP_1}{dP}(x),\ \ x \in R,$$

be corresponding densities. Assume that $p_0(x)$ and $p_1(x)$ are positive and continuous. Denoting $P_n = P^n$, we have

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{dP_{n,\theta}}{dP_n}(X_n) = \frac{1}{n} \sum_{i=1}^{[n\theta]} \log p_0(X_{i,n}) + \frac{1}{n} \sum_{i=[n\theta]+1}^{n} \log p_1(X_{i,n}),$$

so that defining an empirical process again by

$$F_n(x,t) = \frac{1}{n} \sum_{i=1}^{[nt]} 1(X_{i,n} \leq x),\ x \in R, t \in [0,1],$$

we obtain the representation

$$\Xi_{n,\theta} = \int_0^\theta \int_R \log p_0(x)\, F_n(dx, dt) + \int_\theta^1 \int_R \log p_1(x)\, F_n(dx, dt).$$

Let space $\mathcal{Y}$ be defined as for the preceding model and let $\mathcal{Y}_P$ consist of those $F \in \mathcal{Y}$ which are absolutely continuous relative to measure $P(dx) \times dt$ and admit density $p_t(x)$ such that $\int_R p_t(x) P(dx) = 1\,, t \geq 0\,.$ As above, the F_n obey the LDP with rate function I_P^{SK} of the form

$$I_P^{SK}(F) = \begin{cases} \displaystyle\int_0^1 \int_R p_t(x) \log p_t(x) P(dx)\, dt, & \text{if } F \in \mathcal{Y}_P, \\ \infty, & \text{otherwise.} \end{cases}$$

Define next for $F \in \mathcal{Y}_P$

$$\zeta_\theta(F) = \int_0^\theta \int_R \log p_0(x)\, F(dx, dt) + \int_\theta^1 \int_R \log p_1(x)\, F(dx, dt)\,.$$

Then again $\Xi_{n,\theta} = \zeta_\theta(F_n)$ and the LDP for $\{\Xi_{n,\Theta}, n \geq 1\}$ holds, e.g., if $\log p_0(x)$ and $\log p_1(x)$ are bounded and continuous.

Example 2.7 *Regression with Random Design*

We consider the model

$$X_{k,n} = \theta(t_{k,n}) + \xi_{k,n}, \qquad k = 1, \ldots, n,$$

where errors $\xi_{k,n}$ and design points $t_{k,n}$ are independent i.i.d. with respective distributions P, admitting density $p(x)$, and Π. We assume also that the prior measure Π is compactly supported by set D and has positive continuous density $\pi(t)$ on the support. The unknown regression function $\theta(\cdot)$ is assumed to be continuous.

In this model, $P_{n,\theta}$ is the joint distribution of $X_n = (X_{1,n}, \ldots, X_{n,n})$ and $t_n = (t_{1,n}, \ldots, t_{n,n})$ for θ. Let F_n be the joint empirical distribution function of X_n and t_n:

$$F_n(A, B) = \frac{1}{n} \sum_{k=1}^{n} 1(X_{k,n} \in A, t_{k,n} \in B)\,,$$

for Borel sets $A \subset R$, $B \subset D$, and let $\mathcal{Y}$ be the space of probability distributions on $R \times D$ submitted with weak topology. Set also $P_n = P_{n,0} = (P \times \Pi)^n$.

With these definitions,

$$\begin{aligned}
\Xi_{n,\theta} &= \frac{1}{n} \log \frac{dP_{n,\theta}}{dP_n}(X_n, t_n) \\
&= \frac{1}{n} \sum_{k=1}^{n} \log \frac{p(X_{k,n} - \theta(t_{k,n}))}{p(X_{k,n})} \\
&= \int_R \int_R \log \frac{p(x - \theta(t))}{p(x)} F_n(dx, dt).
\end{aligned}$$

Let $\mathcal{Y}_1$ be the set of two-dimensional cumulative distribution functions on R^2 which are absolutely continuous w.r.t. Lebesgue measure on R^2 and have support in $R \times D$. Under P_n, the random pairs $(X_{k,n}, t_{k,n})$ are i.i.d. with distribution $P \times \Pi$, and hence, by Sanov's theorem, the LDP holds for the F_n with the rate function $I^{SS}(F)$ given by

$$I^{SS}(F) = \begin{cases} \displaystyle\int_R \int_D \log \frac{p(x,t)}{p(x)\pi(t)}\, p(x,t)\, dx\, dt, & \text{if } F \in \mathcal{Y}_1, \\ \infty, & \text{otherwise.} \end{cases}$$

Here $F(dx, dt) = p(x,t)\, dx\, dt$.

Further this model can be treated in a manner similar to the case of an i.i.d. sample.

2.2 Sufficient conditions for the dominated LDP

We next study properties of the LDP for statistical experiments and begin with a sufficient condition for the LDP to hold. The condition serves two purposes further:

firstly, in particular statistical models it is easier to be checked than the definition of the LDP; secondly, this condition comes in handy when constructing asymptotically optimal decisions, see Section 4. The idea behind the condition is similar to the one used in the condition of local asymptotic normality by LeCam, 1960, or, more generally, in the condition of λ-convergence by Shiryaev and Spokoiny, 1995, for studying weak convergence of experiments.

Assume that there exist statistics Y_n on $(\Omega_n, \mathcal{F}_n)$ with values in a Hausdorff space $\mathcal{Y}$ such that the sequence $\{\mathcal{L}(Y_n|P_n), n \geq 1\}$ obeys the LDP and the Y_n are asymptotically sufficient in the sense that $Z_{n,\theta} \approx \mathfrak{z}_\theta(Y_n)$ for some nonrandom functions $\mathfrak{z}_\theta$ on $\mathcal{Y}$ (later on we explain the meaning of this approximate equality).

In the above examples, the Y_n are easily identified: it is the empirical mean $(X_{1,n}+\ldots+X_{n,n})/n$ in the case of a sample from normal distribution in Example 2.1, the empirical distribution function F_n in the case of an i.i.d. sample in Example 2.2, the observation process X_n in the "signal + white noise" model, the empirical process F_n in the cases of the regression model with non-Gaussian errors and the change-point model, etc.

If the functions $\mathfrak{z}_\theta$ are bounded and continuous, then, as we have seen, by the contraction principle, the LDP for the sequence $\{\mathcal{L}(Y_n|P_n), n \geq 1\}$ implies the LDP for the sequence $\{\mathcal{L}(\mathfrak{z}_\theta(Y_n)|P_n), n \geq 1\}$ and hence for $\{\mathcal{L}\left(Z_{n,\theta}|P_n\right), n \geq 1\}$. But, by contrast with the theory of weak convergence of experiments, in applications the functions $\mathfrak{z}_\theta$ typically are not continuous. For instance, the functions $\zeta_\theta(y) = \log \mathfrak{z}_\theta(y)$ generally are not continuous in the above examples for an i.i.d. sample, the "signal + white noise" model, the regression models, the change-point model. To cope with this, we invoke the idea of regularisation which makes the condition more complicated.

For the sequel, we need some more definitions and facts from large deviation theory. Recall, see Varadhan, 1966; Varadhan, 1984; Deuschel and Stroock, 1989; Bryc, 1990, that if a sequence of probability measures $\{Q_n, n \geq 1\}$ on the Borel σ-field of a Hausdorff space S obeys the LDP with rate function I, then, for all nonnegative, bounded, continuous functions f on S,

$$\lim_{n\to\infty} \left[\int_S (f(x))^n Q_n(dx)\right]^{1/n} = \sup_{x\in S} f(x)V(x), \tag{2.5}$$

where $V(x) = \exp(-I(x))$. If S is a metric, or more generally, a Tihonov (i.e., completely regular) space, then (2.5) is also sufficient for the LDP, see Puhalskii, 1993.

Moreover, the LDP implies (2.5) for unbounded continuous nonnegative functions f too under "the uniform exponential integrability condition", Varadhan, 1984; Deuschel and Stroock, 1989,

$$\lim_{H\to\infty} \overline{\lim_{n\to\infty}} \left[\int_S (f(x))^n \, 1(f(x) > H) Q_n(dx)\right]^{1/n} = 0. \tag{2.6}$$

Also, if f is a lower semicontinuous nonnegative function, then

$$\underline{\lim}_{n\to\infty} \left[\int_S (f(x))^n Q_n(dx)\right]^{1/n} \geq \sup_{x\in S} f(x)V(x).$$

The function $V(x)$ is further referred to as deviability. Equivalently, a deviability is defined as a function $V : S \to [0,1]$ such that $\sup_{x\in S} V(x) = 1$ and the sets $V^{-1}[a,1]$ are compact for all $a > 0$. Obviously, there is one-to-one correspondence between probability rate functions and deviabilities. We will be saying that $\{Q_n, n \geq 1\}$ large deviation (LD) converges to V and write $Q_n \stackrel{l.d.}{\to} V$ $(n \to \infty)$ if (2.5) holds for all bounded continuous nonnegative functions f (Puhalskii, 1994a). Below, we will be using the fact that, if S is metric, one can require that the functions f be uniformly continuous (analogously to weak convergence theory, Billingsley, 1968, Theorem 2.1). By the above, if S is a Tihonov space, then $Q_n \stackrel{l.d.}{\to} V$ $(n \to \infty)$ if and only if $\{Q_n\}$ obeys the LDP with $I = -\log V$. All the spaces we are considering below are Tihonov and we will mostly be using the formulation of the LDP as LD convergence as more convenient in theoretical considerations.

Next, let S and S' be Hausdorff spaces, and let V be a deviability on S. Denote

$$\Phi_V(a) = \{x \in S : V(x) \geq a\}, \quad a > 0. \tag{2.7}$$

As in Puhalskii, 1995b (cf. Schwartz, 1973), we will say that a map $\varphi : S \to S'$ is V-Luzin measurable if it is continuous in restriction to each set $\Phi_V(a), a > 0$. Deviabilities are preserved under Luzin measurable maps: for any V-Luzin measurable map φ, the function $V \circ \varphi^{-1}$ on S' defined by $V \circ \varphi^{-1}(x') = \sup_{x\in\varphi^{-1}(x')} V(x), \quad x' \in S'$, is a deviability on S' (e.g., the argument of Puhalskii, 1991, Lemma 2.1 applies).

Further, say that $\varphi : S \to S'$ is V–almost everywhere (V-a.e.) continuous if it is continuous at any $x \in S$ such that $V(x) > 0$. Obviously, any V-a.e. continuous function is V-Luzin measurable.

We introduce more notational conventions. $\mathcal{A}(\Theta)$ denotes further the family of all finite subsets of Θ. The elements of R_+^Θ are denoted by $z_\Theta = (z_\theta, \theta \in \Theta)$, and the elements of R_+^Λ, where $\Lambda \in \mathcal{A}(\Theta)$, are denoted by $z_\Lambda = (z_\theta, \theta \in \Lambda)$. Maps π_Λ and $\pi_{\Lambda'\Lambda}$, where $\Lambda \in \mathcal{A}(\Theta)$, $\Lambda' \in \mathcal{A}(\Theta)$ and $\Lambda \subset \Lambda'$, are natural projections of R_+^Θ onto R_+^Λ and of $R_+^{\Lambda'}$ onto R_+^Λ, respectively: $\pi_\Lambda(z_\theta, \theta \in \Theta) = (z_\theta, \theta \in \Lambda)$ and $\pi_{\Lambda'\Lambda}(z_\theta, \theta \in \Lambda') = (z_\theta, \theta \in \Lambda)$. Since R_+^Θ and R_+^Λ, $\Lambda \in \mathcal{A}(\Theta)$, are submitted with Tihonov topology, the projections are continuous.

We now state and prove the sufficient condition for the LDP. In it we assume that the statistics Y_n take values in a metric space which is enough for applications though this restriction can be relaxed.

Lemma 2.1 *Let $\{\mathcal{E}_n, P_n, n \geq 1\}$ be a sequence of dominated experiments and let $Z_{n,\theta}$, $\theta \in \Theta$, be defined by (2.1).*

Assume that the following condition holds:

(Y) there exist statistics $Y_n : \Omega_n \to \mathcal{Y}$ with values in a metric space $\mathcal{Y}$ with Borel σ-field, functions $\mathfrak{z}_\theta : \mathcal{Y} \to R_+$, $\theta \in \Theta$ and $\mathfrak{z}_{\theta,\delta} : \mathcal{Y} \to R_+$, $\theta \in \Theta$, $\delta > 0$, such that

(Y.1) *the sequence $\{\mathcal{L}(Y_n|P_n), n \geq 1\}$ of distributions on $\mathcal{Y}$ LD converges to deviability $V(y), y \in \mathcal{Y}$;*

(Y.2) *for each $\delta > 0$, the functions $\mathfrak{z}_{\theta,\delta} : \mathcal{Y} \to R_+$, $\theta \in \Theta$, are Borel and V-a.e. continuous;*

$$(Y.3) \qquad \lim_{\delta\to 0} \overline{\lim_{n\to\infty}} P_n^{1/n}(|Z_{n,\theta} - \mathfrak{z}_{\theta,\delta}(Y_n)| > \varepsilon) = 0 \textit{ for all } \varepsilon > 0 \textit{ and } \theta \in \Theta;$$

$$(Y.4) \qquad \lim_{\delta\to 0} \sup_{y\in\Phi_V(a)} |\mathfrak{z}_{\theta,\delta}(y) - \mathfrak{z}_\theta(y)| = 0 \textit{ for all } a > 0 \textit{ and } \theta \in \Theta.$$

Then $\mathcal{L}(Z_{n,\Theta}|P_n) \overset{l.d.}{\to} V_\Theta$ $(n\to\infty)$, *where* $V_\Theta = V\circ \mathfrak{z}_\Theta^{-1}$, $\mathfrak{z}_\Theta = (\mathfrak{z}_\theta,\, \theta\in\Theta)$.

Proof Conditions $(Y.2)$ and $(Y.4)$ obviously imply that $\mathfrak{z}_\Theta : \mathcal{Y} \to R_+^\Theta$ is V-Luzin measurable, hence V_Θ is a deviability on R_+^Θ.

Let $\Lambda \in \mathcal{A}(\Theta)$. We first prove that

$$\mathcal{L}(Z_{n,\Lambda}|P_n) \overset{l.d.}{\to} V_\Lambda \;\; (n\to\infty), \tag{2.8}$$

where $Z_{n,\Lambda} = (Z_{n,\theta}, \theta\in\Lambda)$, $V_\Lambda = V\circ\mathfrak{z}_\Lambda^{-1}$ and $\mathfrak{z}_\Lambda = (\mathfrak{z}_\theta,\, \theta\in\Lambda)$. Let $f : R_+^\Lambda \to R_+$ be bounded and uniformly continuous. Since, by the definition of V_Λ, $\sup_{z_\Lambda\in R_+^\Lambda} f(z_\Lambda)V_\Lambda(z_\Lambda) = \sup_{y\in\mathcal{Y}} f(\mathfrak{z}_\Lambda(y))V(y)$, we need to prove that

$$\lim_{n\to\infty} E_n^{1/n} f^n(Z_{n,\Lambda}) = \sup_{y\in\mathcal{Y}} f(\mathfrak{z}_\Lambda(y))V(y). \tag{2.9}$$

Condition $(Y.3)$ implies in view of the boundedness and uniform continuity of f that

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}} |E_n^{1/n} f^n(Z_{n,\Lambda}) - E_n^{1/n} f^n(\mathfrak{z}_{\Lambda,\delta}(Y_n))| = 0, \tag{2.10}$$

where $\mathfrak{z}_{\Lambda,\delta} = (\mathfrak{z}_{\theta,\delta}, \theta\in\Lambda)$.

Since the sequence $\{\mathcal{L}(Y_n|P_n), n\geq 1\}$ LD converges to V and $\mathfrak{z}_{\Lambda,\delta} : \mathcal{Y}\to R_+^\Lambda$ are V-a.e. continuous, the sequence $\{\mathcal{L}(\mathfrak{z}_{\Lambda,\delta}(Y_n)|P_n), n\geq 1\}$ LD converges to $V\circ(\mathfrak{z}_{\Lambda,\delta})^{-1}$, Puhalskii, 1991. Thus, since f is bounded and continuous,

$$\lim_{n\to\infty} E_n^{1/n} f^n(\mathfrak{z}_{\Lambda,\delta}(Y_n)) = \sup_{y\in\mathcal{Y}} f(\mathfrak{z}_{\Lambda,\delta}(y))V(y). \tag{2.11}$$

Due to (2.10) and (2.11), for (2.9) it remains to show that

$$\lim_{\delta\to 0}\sup_{y\in\mathcal{Y}} f(\mathfrak{z}_{\Lambda,\delta}(y))V(y) = \sup_{y\in\mathcal{Y}} f(\mathfrak{z}_\Lambda(y))V(y), \tag{2.12}$$

which is an easy consequence of condition $(Y.4)$. Convergence (2.8) is proved. The assertion of the lemma now follows by Dawson–Gärtner's theorem on projective limits of large deviation systems, Dawson and Gärtner, 1987, if we note that $\mathcal{L}(Z_{n,\Theta}|P_n)$ is the projective limit of $\{\mathcal{L}(Z_{n,\Lambda}|P_n), \Lambda\in\mathcal{A}(\Theta)\}$ and V_Θ is the projective limit of $\{V_\Lambda, \Lambda\in\mathcal{A}(\Theta)\}$, the latter meaning that the corresponding rate function I_Θ is the projective limit of $\{I_\Lambda, \Lambda\in\mathcal{A}(\Theta)\}$. □

Remark 2.1 *Since* R_+^Θ *is a Tihonov space, according to the lemma, the sequence* $\{\mathcal{E}_n, P_n, n\geq 1\}$ *obeys the dominated LDP if conditions* (Y) *and* (U) *hold.*

Remark 2.2 *As we have seen, in applications it is more convenient to manipulate rate functions and the log-likelihood ratios given by*

$$\Xi_{n,\theta} = \log Z_{n,\theta} = \frac{1}{n} \log \frac{dP_{n,\theta}}{dP_n}, \quad \theta \in \Theta.$$

Assuming that the $\Xi_{n,\theta}$ are well-defined, condition (Y) is implied by the following condition

(Y') *there exist statistics $Y_n : \Omega_n \to \mathcal{Y}$ with values in a metric space $\mathcal{Y}$ with Borel σ-field, functions $\zeta_\theta : \mathcal{Y} \to R$, $\theta \in \Theta$, and $\zeta_{\theta,\delta} : \mathcal{Y} \to R$, $\theta \in \Theta$, $\delta > 0$, such that*

$(Y'.1)$ *the sequence $\{\mathcal{L}(Y_n | P_n),\, n \geq 1\}$ of distributions on $\mathcal{Y}$ satisfies the LDP on $\mathcal{Y}$ with rate function $I(y)$, $y \in \mathcal{Y}$;*

$(Y'.2)$ *for each $\delta > 0$, the functions $\zeta_{\theta,\delta} : \mathcal{Y} \to R$, $\theta \in \Theta$, are Borel and continuous at each point y such that $I(y) < \infty$;*

$(Y'.3)$ $\quad \lim\limits_{\delta \to 0} \overline{\lim\limits_{n\to\infty}} \; P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(Y_n)| > \varepsilon) = 0$ *for all $\varepsilon > 0$ and $\theta \in \Theta$;*

$(Y'.4)$ $\quad \lim\limits_{\delta \to 0} \sup\limits_{y \in \Phi_I'(a)} |\zeta_{\theta,\delta}(y) - \zeta_\theta(y)| = 0$ *for all $a \geq 0$ and $\theta \in \Theta$,*

where $\Phi_I'(a) = \{y \in \mathcal{Y} : I(y) \leq a\}$.

Condition (U) takes the form

(U') $$\lim_{H\to\infty} \overline{\lim_{n\to\infty}} \; E_n^{1/n} \exp(n\Xi_{n,\theta}) 1(\Xi_{n,\theta} > H) = 0, \;\; \theta \in \Theta\,.$$

By Lemma 2.1, conditions (Y') and (U') imply the dominated LDP.

2.3 The general case

The above definition of the large deviation principle for statistical experiments covers only the dominated case and depends on the choice of dominating measures. We present now another definition which is free of these defects. It is motivated by Le Cam's definition of weak convergence of experiments, see, e.g., Strasser, 1985.

Let $|\Lambda|$ denote the number of elements in $\Lambda \in \mathcal{A}(\Theta)$. For $z_\Lambda = (z_\theta, \theta \in \Lambda) \in R_+^\Lambda$ and $z_\Theta = (z_\theta, \theta \in \Theta) \in R_+^\Theta$, we set $\|z_\Lambda\|_\Lambda = \max\limits_{\theta\in\Lambda} z_\theta$ and $\|z_\Theta\|_\Theta = \max\limits_{\theta\in\Theta} z_\theta$, respectively, and define $S_\Lambda = \{z_\Lambda \in R_+^\Lambda : \|z_\Lambda\|_\Lambda = 1\}$ and $S_\Theta = \{z_\Theta \in R_+^\Theta : \|z_\Theta\|_\Theta = 1\}$. Not to overburden notation, we sometimes omit subscript Λ in $\|\cdot\|_\Lambda$ if there is no risk of confusion.

Next, given a sequence of experiments $\{\mathcal{E}_n\,, n \geq 1\}$, set, for each $\Lambda \in \mathcal{A}(\Theta)$,

$$\begin{aligned} P_{n,\Lambda} &= \frac{1}{|\Lambda|} \sum_{\theta\in\Lambda} P_{n,\theta}, \\ \mathbf{Z}_{n,\theta;\Lambda} &= \left(\frac{dP_{n,\theta}}{dP_{n,\Lambda}}\right)^{1/n}, \qquad \theta \in \Lambda, \qquad\qquad (2.13) \\ \mathbf{Z}_{n,\Lambda} &= (\mathbf{Z}_{n,\theta;\Lambda},\, \theta \in \Lambda). \end{aligned}$$

The definitions immediately imply that, $P_{n,\Lambda}$-a.s.,

$$\sum_{\theta\in\Lambda} \mathbf{Z}^n_{n,\theta;\Lambda} = |\Lambda| \tag{2.14}$$

and

$$1 \le \|\mathbf{Z}_{n,\Lambda}\| \le |\Lambda|^{1/n}. \tag{2.15}$$

Definition 2.2 *A sequence* $\{\mathcal{E}_n\,, n \ge 1\}$ *of statistical experiments obeys the LDP if, for each* $\Lambda \in \mathcal{A}(\Theta)$*, the sequence of distributions* $\{\mathcal{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}),\, n \ge 1\}$ *obeys the LDP on* R_+^Λ *with some rate function.*

Remark 2.3 *Equivalently,* $\{\mathcal{E}_n\,, n \ge 1\}$ *obeys the LDP if* $\mathcal{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \overset{l.d.}{\to} \mathbf{V}_\Lambda$, $\Lambda \in \mathcal{A}(\Theta)$*, where* $\mathbf{V}_\Lambda$ *is a deviability on* R_+^Λ*.*

We next study consequences of the definition and prove, in particular, that the definitions of the LDP for the dominated and general cases are consistent.

Lemma 2.2 *Let* $\Lambda \in \mathcal{A}(\Theta)$*. If* $\mathcal{L}\,(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \overset{l.d.}{\to} \mathbf{V}_\Lambda$*, then* $\mathbf{V}_\Lambda$ *has support in* S_Λ*, i.e.,* $\mathbf{V}_\Lambda(z_\Lambda) = 0$ *if* $z_\Lambda \notin S_\Lambda$*.*

Proof We have using the equivalence of LD convergence and LDP on R_+^Λ, that, for $\varepsilon > 0$,

$$\underline{\lim}_{n\to\infty}\; P_{n,\Lambda}^{1/n}(|\|\mathbf{Z}_{n,\Lambda}\| - 1| > \varepsilon) \ge \sup_{z_\Lambda:\, |\|z_\Lambda\|-1|>\varepsilon} \mathbf{V}_\Lambda(z_\Lambda).$$

Inequalities (2.15) imply that the left hand side is zero. Since ε is arbitrary, $\mathbf{V}_\Lambda(z_\Lambda) = 0$ if $\|z_\Lambda\| \ne 1$. $\square$

We now give another characterisation of the LDP. Let $\mathcal{H}_\Lambda$ denote the set of all nonnegative, continuous and positively homogeneous functions on R_+^Λ: $h \in \mathcal{H}_\Lambda$ iff $h(z_\Lambda) \ge 0$, h is continuous and $h(\lambda z_\Lambda) = \lambda h(z_\Lambda)$ for all $z_\Lambda \in R_+^\Lambda$ and $\lambda \ge 0$.

Lemma 2.3 *Let* $\Lambda \in \mathcal{A}(\Theta)$*. Then* $\mathcal{L}\,(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \overset{l.d.}{\to} \mathbf{V}_\Lambda$ *if and only if* $\mathbf{V}_\Lambda$ *has support in* S_Λ *and*

$$\lim_{n\to\infty} E_{n,\Lambda}^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) = \sup_{z_\Lambda\in R_+^\Lambda} h(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda)\,, \quad \textit{for any } h \in \mathcal{H}_\Lambda\,.$$

Proof Let $\mathcal{L}\,(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \overset{l.d.}{\to} \mathbf{V}_\Lambda$. Then $\mathbf{V}_\Lambda$ has support in S_λ by Lemma 2.2. The second claim follows by the definition of LD convergence since, by (2.15), $h(\mathbf{Z}_{n,\Lambda}) = \hat{h}(\mathbf{Z}_{n,\Lambda})$ $P_{n,\Lambda}$-a.s., where $\hat{h}(z_\Lambda) = h(z_\Lambda)[(2 - \|z_\Lambda\|/\Lambda) \wedge 1 \vee 0]$, and the latter function is bounded and continuous.

For the converse, pick a nonnegative continuous bounded function f on R_+^Λ. We need to prove that

$$\lim_{n\to\infty} E_{n,\Lambda}^{1/n} f^n(\mathbf{Z}_{n,\Lambda}) = \sup_{z_\Lambda\in R_+^\Lambda} f(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda). \tag{2.16}$$

We define the function $\tilde f$ by

$$\tilde f(z_\Lambda) = \begin{cases} \|z_\Lambda\| f\left(\dfrac{z_\Lambda}{\|z_\Lambda\|}\right), & \|z_\Lambda\| > 0\,, \\ 0\,, & \|z_\Lambda\| = 0\,. \end{cases}$$

Note that f and $\tilde f$ coincide on S_Λ and, since $\mathbf{V}_\Lambda$ is supported by S_Λ, we can change f to $\tilde f$ on the right hand-side of (2.16). The continuity of f and the inequalities (2.15) easily imply that the random variables $f(\mathbf{Z}_{n,\Lambda})$ and $\tilde f(\mathbf{Z}_{n,\Lambda})$ are uniformly bounded and

$$\lim_{n\to\infty} \left| E_{n,\Lambda}^{1/n} f^n(\mathbf{Z}_{n,\Lambda}) - E_{n,\Lambda}^{1/n} \tilde f^n(\mathbf{Z}_{n,\Lambda}) \right| = 0.$$

Since $\tilde f \in \mathcal{H}_\Lambda$, taking $h = \tilde f$ in the conditions of the lemma, we get

$$\lim_{n\to\infty} E_{n,\Lambda}^{1/n} \tilde f^n(\mathbf{Z}_{n,\Lambda}) = \sup_{z_\Lambda \in R_+^\Lambda} \tilde f(z_\Lambda) \mathbf{V}_\Lambda(z_\Lambda),$$

which yields (2.16) as required. □

We now show that if $\Lambda \subset \Lambda' \in \mathcal{A}(\Theta)$, then deviability $\mathbf{V}_\Lambda$ is a sort of projection of deviability $\mathbf{V}_{\Lambda'}$, the property being inherited from corresponding probabilities. Recall the notations $\pi_{\Lambda'\Lambda}$ and π_Λ for projections from $R_+^{\Lambda'}$ onto R_+^Λ and R_+^Θ onto R_+^Λ, respectively, and let $\Pi_{\Lambda'\Lambda}$ and Π_Λ stand for normalised projections:

$$\Pi_{\Lambda'\Lambda} z_{\Lambda'} = \pi_{\Lambda'\Lambda} z_{\Lambda'} / \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda, \quad z_{\Lambda'} \in R_+^{\Lambda'}\,, \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda > 0\,,$$
$$\Pi_\Lambda z_\Theta = \pi_\Lambda z_\Theta / \|\pi_\Lambda z_\Theta\|_\Lambda, \quad z_\Theta \in R_+^\Theta\,, \|\pi_\Lambda z_\Theta\|_\Lambda > 0\,.$$

Lemma 2.4 *Let* $\Lambda \subset \Lambda' \in \mathcal{A}(\Theta)$. *If* $\mathcal{L}(\mathbf{Z}_{n,\Lambda} | P_{n,\Lambda}) \stackrel{l.d.}{\to} \mathbf{V}_\Lambda$ *and* $\mathcal{L}(\mathbf{Z}_{n,\Lambda'} | P_{n,\Lambda'}) \stackrel{l.d.}{\to} \mathbf{V}_{\Lambda'}$, *then the following conditions hold*

$$(C) \qquad \sup_{z_\Lambda \in R_+^\Lambda} h(z_\Lambda) \mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_{\Lambda'} \in R_+^{\Lambda'}} h(\pi_{\Lambda'\Lambda} z_{\Lambda'}) \mathbf{V}_{\Lambda'}(z_{\Lambda'})\,,\ h \in \mathcal{H}_\Lambda\,;$$

$$(S) \qquad \mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_{\Lambda'} \in \Pi_{\Lambda'\Lambda}^{-1} z_\Lambda} \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}),\ z_\Lambda \in R_+^\Lambda\,,$$

where $\Pi_{\Lambda'\Lambda}^{-1} z_\Lambda = \{z_{\Lambda'} \in R_+^{\Lambda'} \,:\, \Pi_{\Lambda'\Lambda} z_{\Lambda'} = z_\Lambda\}$.

Proof Define

$$\mathbf{Z}_{n,\Lambda;\Lambda'} = \left(\frac{dP_{n,\Lambda}}{dP_{n,\Lambda'}}\right)^{1/n}.$$

Then obviously

$$\pi_{\Lambda'\Lambda} \mathbf{Z}_{n,\Lambda'} = \mathbf{Z}_{n,\Lambda} \mathbf{Z}_{n,\Lambda;\Lambda'} \qquad P_{n,\Lambda'}\text{– a.s.},$$

and, since $h \in \mathcal{H}_\Lambda$, we have that

$$E_{n,\Lambda}^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) = E_{n,\Lambda'}^{1/n} \left[h(\mathbf{Z}_{n,\Lambda'}) \mathbf{Z}_{n,\Lambda;\Lambda'}\right]^n = E_{n,\Lambda'}^{1/n} h^n(\pi_{\Lambda',\Lambda} \mathbf{Z}_{n,\Lambda'}).$$

Applying Lemma 2.3 to the leftmost and rightmost sides and using that $h \circ \pi_{\Lambda'\Lambda} \in \mathcal{H}_{\Lambda'}$, we obtain (C).

Now, (S), for given $\hat{z}_\Lambda \in S_\Lambda$, can formally be obtained by substituting $h(z_\Lambda) = \mathbf{1}(z_\Lambda = \|z_\Lambda\|\hat{z}_\Lambda)\|z_\Lambda\|$ in (C) and using that $\mathbf{V}_\Lambda$ has support in S_Λ. However the function h is not continuous, so we approximate it by a sequence of continuous functions $h_k \in \mathcal{H}_\Lambda$, $k \geq 1$, as follows. Let

$$h_k(z_\Lambda) = (\|z_\Lambda\| - k\|z_\Lambda - \hat{z}_\Lambda\|z_\Lambda\|\|)^+.$$

Since the h_k are from $\mathcal{H}_\Lambda$, we can apply (C). Also $h_k(z_\Lambda) \downarrow h(z_\Lambda)$ as $k \to \infty$. Using that $h(z_\Lambda)$ is upper semicontinuous, and $\mathbf{V}_\Lambda$ and $\mathbf{V}_{\Lambda'}$ are deviabilities, it is not difficult to see (see also Puhalskii, 1995b) that one can take limit as $k \to \infty$ in (C) for the h_k's, as required. $\square$

Remark 2.4 *We have actually proved that (C) holds for non continuous positively homogeneous nonnegative functions too.*

We further call a family of deviabilities $\{V_\Lambda, \Lambda \in \mathcal{A}(\Theta)\}$, where V_Λ is defined on R_+^Λ, *conical* if it satisfies (C). If, in addition, $V_\Lambda(z_\Lambda) = 0$ for all $z_\Lambda \notin S_\Lambda$, the family is called *standard*. By the above, a family is standard if it meets (S).

The next result is of particular importance for the minimax theorem below. It states that any standard family of deviabilities admits an extension on R_+^Θ which preserves the conical property.

Lemma 2.5 *For any standard family of deviabilities $\{\mathbf{V}_\Lambda, \Lambda \in \mathcal{A}(\Theta)\}$, there exists a function $\mathbf{V}_\Theta$ on R_+^Θ such that the following conditions hold:*

(i) $\mathbf{V}_\Theta$ is upper semicontinuous on R_+^Θ, assumes values in $[0,1]$, $\sup_{z_\Theta \in R_+^\Theta} \mathbf{V}_\Theta(z_\Theta) = 1$ and $\mathbf{V}_\Theta(z_\Theta) = 0$ if $z_\Theta \notin S_\Theta$;

(ii) for any $\Lambda \in \mathcal{A}(\Theta)$ and $h \in \mathcal{H}_\Lambda$,

$$\sup_{z_\Lambda \in R_+^\Lambda} h(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta \in R_+^\Theta} h(\pi_\Lambda z_\Theta)\mathbf{V}_\Theta(z_\Theta);$$

(iii) for any $z_\Lambda \in R_+^\Lambda$,

$$\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta \in \Pi_\Lambda^{-1} z_\Lambda} \|\pi_\Lambda z_\Theta\|_\Lambda \mathbf{V}_\Theta(z_\Theta),$$

where $\Pi_\Lambda^{-1} z_\Lambda = \{z_\Theta \in R_+^\Theta : \Pi_\Lambda z_\Theta = z_\Lambda\}$.

A proof is deferred to the appendix.

We conclude the section by showing the consistency of the above definitions of the LDP.

Lemma 2.6 *Let $\{\mathcal{E}_n, P_n, n \geq 1\}$ be a sequence of dominated statistical experiments. If $\{\mathcal{E}_n, P_n, n \geq 1\}$ obeys the dominated LDP, then it obeys the LDP. More specifically, if deviability V_Θ on R_+^Θ is the LD limit of $\mathcal{L}(Z_{n,\Theta}|P_n)$ as $n \to \infty$, then $\mathcal{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \overset{l.d.}{\to} \mathbf{V}_\Lambda$, $\Lambda \in \mathcal{A}(\Theta)$, where*

$$\mathbf{V}_\Lambda(z_\Lambda) = \begin{cases} \sup_{z_\Theta \in \Pi_\Lambda^{-1} z_\Lambda} \|\pi_\Lambda z_\Theta\| V_\Theta(z_\Theta), & z_\Lambda \in S_\Lambda, \\ 0, & z_\Lambda \notin S_\Lambda. \end{cases}$$

Proof Let $\mathcal{L}(Z_{n,\Theta}|P_n) \stackrel{l.d.}{\to} V_\Theta$. It is easy to see that the $\mathbf{V}_\Lambda$ defined in the statement are deviabilities with support in S_Λ.

By Lemma 2.3, it suffices then to prove that, for any $\Lambda \in \mathcal{A}(\Theta)$ and $h \in \mathcal{H}_\Lambda$,

$$\lim_{n\to\infty} E_{n,\Lambda}^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) = \sup_{z_\Lambda \in R_+^\Lambda} h(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda).$$

Denote

$$\alpha_{n,\Lambda} = \left(\frac{dP_{n,\Lambda}}{dP_n}\right)^{1/n}.$$

We have, in earlier notation,

$$\alpha_{n,\Lambda}^n = \frac{dP_{n,\Lambda}}{dP_n} = \frac{1}{|\Lambda|}\sum_{\theta\in\Lambda}\frac{dP_{n,\theta}}{dP_n} = \frac{1}{|\Lambda|}\sum_{\theta\in\Lambda} Z_{n,\theta}^n,$$

and

$$Z_{n,\Lambda} = \mathbf{Z}_{n,\Lambda}\alpha_{n,\Lambda} \qquad P_n\text{–a.s.}$$

Hence

$$E_{n,\Lambda}^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) = E_n^{1/n} h^n(\mathbf{Z}_{n,\Lambda})\alpha_{n,\Lambda}^n = E_n^{1/n} h^n(Z_{n,\Lambda}). \tag{2.17}$$

Now we are using the LD convergence $L(Z_{n,\Theta}|P_n) \stackrel{l.d.}{\to} V_\Theta$. However, we cannot apply at this point property (2.5) to the function $h(z_\Lambda)$ since it is not bounded on R_+^Λ. So we check (2.6). This is where condition (U) comes in.

Let $h^* = \sup_{z_\Lambda\in S_\Lambda} h(z_\Lambda)$. Since h is continuous, it is bounded on S_Λ and $h^* < \infty$. Since $h \in \mathcal{H}_\Lambda$, $h(Z_{n,\Lambda}) \le h^*\|Z_{n,\Lambda}\|$ and, in view of condition (U),

$$\begin{aligned}\overline{\lim_{n\to\infty}}\ E_n^{1/n} h^n(Z_{n,\Lambda})1(h(Z_{n,\Lambda}) > H) &\le \overline{\lim_{n\to\infty}} \sum_{\theta\in\Lambda} E_n^{1/n} h^{*n} Z_{n,\theta}^n 1(h^* Z_{n,\theta} > H)\\ &\le \overline{\lim_{n\to\infty}}\ h^* \sum_{\theta\in\Lambda} P_{n,\theta}^{1/n}(h^* Z_{n,\theta} > H) \to 0 \text{ as } H\to\infty.\end{aligned}$$

Property (2.6) is checked and we obtain, by (2.17), (2.5) and the LD convergence of $\mathcal{L}(Z_{n,\Theta}|P_n)$ to V_Θ, that

$$\begin{aligned}\lim_{n\to\infty} E_{n,\Lambda}^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) &= \lim_{n\to\infty} E_n^{1/n} h^n(Z_{n,\Lambda})\\ &= \lim_{n\to\infty} E_n^{1/n} h^n(\pi_\Lambda Z_{n,\Theta})\\ &= \sup_{z_\Theta\in R_+^\Theta} h(\pi_\Lambda z_\Theta)V_\Theta(z_\Theta).\end{aligned}$$

Since the definition of $\mathbf{V}_\Lambda$ obviously implies that

$$\sup_{z_\Lambda\in R_+^\Lambda} h(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta\in R_+^\Theta} h(\pi_\Lambda z_\Theta)V_\Theta(z_\Theta),$$

the lemma is proved. □

3 A Minimax Theorem

We start the section by showing that, in analogy with the classical asymptotic theory of statistical experiments, see Strasser, 1985, the LDP allows us to obtain asymptotic lower bounds for appropriately defined risks which, in fact, has been the motivation for introducing the concept of the LDP for sequences of statistical experiments. In the second part of the section we show that, under additional conditions, the bounds are tight and study the problem of constructing sequences of decisions attaining the bounds.

We consider a sequence $\{\mathcal{E}_n\,, n \geq 1\}$, where $\mathcal{E}_n = (\Omega_n, \mathcal{F}_n; P_{n,\theta},\ \theta \in \Theta)$, of statistical experiments and assume that it satisfies the LDP. The corresponding deviabilities are denoted by $\mathbf{V}_\Lambda\,, \Lambda \in \mathcal{A}(\Theta)\,$, and $\mathbf{V}_\Theta$ denotes the extension defined in Lemma 2.5.

We introduce some more notation common for statistical decision theory, see, e.g., Strasser, 1985. We denote by $\mathcal{D}$ a Hausdorff topological space with Borel σ-field which we take as a decision space; $W_\theta = (W_\theta(r), r \in \mathcal{D})$, $\theta \in \Theta$, are, for each θ, nonnegative and lower semicontinuous functions on $\mathcal{D}$ which play the part of loss functions. $\mathcal{R}_n$ denotes the set of all measurable mappings $\rho_n : \Omega_n \to \mathcal{D}$, i.e., $\mathcal{R}_n$ is the set of all decision functions with values in $\mathcal{D}$. We define the large deviation (LD) risk of a decision $\rho_n \in \mathcal{R}_n$ in the experiment $\mathcal{E}_n$ by

$$R_n(\rho_n) = \sup_{\theta\in\Theta} E_{n,\theta}^{1/n} W_\theta^n(\rho_n). \tag{3.1}$$

Obviously, this is an analogue of the risk in minimax decision theory, cf. Strasser, 1985.

Recall, Strasser, 1985, Definition 6.3, that a function $f : U \to R$ on a topological space U is level compact if it is bounded below and the sets $\{u \in U : f(u) \leq \alpha\}$ are compact for all $\alpha < \sup_{u\in U} f(u)$. Obviously, if U is Hausdorff, a level compact function is lower semicontinuous.

Theorem 3.1 *Let the sequence $\{\mathcal{E}_n\,, n \geq 1\}$ obey the LDP. Assume that the functions $W_\theta, \theta \in \Theta$, are level compact. Then*

$$\varliminf_{n\to\infty} \inf_{\rho_n\in\mathcal{R}_n} R_n(\rho_n) \geq R^*,$$

where

$$R^* = \sup_{z_\Theta\in R_+^\Theta} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta} W_\theta(r) z_\theta \mathbf{V}_\Theta(z_\Theta).$$

In particular, if $\{\mathcal{E}_n, P_n, n \geq 1\}$ obeys the dominated LDP and V_Θ is the corresponding deviability, then

$$R^* = \sup_{z_\Theta\in R_+^\Theta} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta} W_\theta(r) z_\theta V_\Theta(z_\Theta). \tag{3.2}$$

If, moreover, conditions (Y) and (U) hold, then

$$R^* = \sup_{y\in\mathcal{Y}} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta} W_\theta(r) \mathfrak{z}_\theta(y) V(y).$$

Proof Let $\Lambda \in \mathcal{A}(\Theta)$. We prove first that

$$\varliminf_{n\to\infty} \inf_{\rho_n} \sup_{\theta\in\Lambda} E_{n,\theta}^{1/n} W_\theta^n(\rho_n) \geq \sup_{z_\Lambda \in R_+^\Lambda} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta \mathbf{V}_\Lambda(z_\Lambda). \tag{3.3}$$

Let $\{\rho_n, n\geq 1\}$ be an arbitrary sequence of decisions. We have, by the definition of $\mathbf{Z}_{n,\Lambda}$ (see (2.13)), that

$$\begin{aligned}
\varliminf_{n\to\infty} \sup_{\theta\in\Lambda} E_{n,\theta}^{1/n} W_\theta^n(\rho_n) &= \varliminf_{n\to\infty} \sup_{\theta\in\Lambda} E_{n,\Lambda}^{1/n} W_\theta^n(\rho_n) \mathbf{Z}_{n,\theta;\Lambda}^n \\
\geq \varliminf_{n\to\infty} \left[\frac{1}{|\Lambda|} E_{n,\Lambda} \sum_{\theta\in\Lambda} W_\theta^n(\rho_n) \mathbf{Z}_{n,\theta;\Lambda}^n\right]^{1/n} &\geq \varliminf_{n\to\infty} E_{n,\Lambda}^{1/n} \sup_{\theta\in\Lambda} W_\theta^n(\rho_n) \mathbf{Z}_{n,\theta;\Lambda}^n \\
&\geq \varliminf_{n\to\infty} E_{n,\Lambda}^{1/n} w^n(\mathbf{Z}_{n,\Lambda}),
\end{aligned}$$

where

$$w(z_\Lambda) = \inf_{r\in\mathcal{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta, \quad z_\Lambda = (z_\theta, \theta\in\Lambda) \in R_+^\Lambda.$$

Since the set Λ is finite and the functions W_θ are lower semicontinuous and level compact, the function $w(\cdot)$ is lower semicontinuous (cf. Aubin, 1984, Proposition 1.7). So by the LD convergence of $\mathcal{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda})$ to $\mathbf{V}_\Lambda$,

$$\varliminf_{n\to\infty} E_{n,\Lambda}^{1/n} w^n(\mathbf{Z}_{n,\Lambda}) \geq \sup_{z_\Lambda\in R_+^\Lambda} w(z_\Lambda) \mathbf{V}_\Lambda(z_\Lambda),$$

implying (3.3).

Since the function $w(\cdot)$ belongs to $\mathcal{H}_\Lambda$, an application of Lemma 2.5(ii) yields,

$$\sup_{z_\Lambda\in R_+^\Lambda} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta \mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta\in R_+^\Theta} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta \mathbf{V}_\Theta(z_\Theta),$$

so by (3.3),

$$\varliminf_{n\to\infty} \inf_{\rho_n} \sup_{\theta\in\Lambda} E_{n,\theta}^{1/n} W_\theta^n(\rho_n) \geq \sup_{z_\Theta\in R_+^\Theta} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta \mathbf{V}_\Theta(z_\Theta).$$

Now the proof is completed by observing that, for every $z_\Theta = (z_\theta, \theta\in\Theta) \in R_+^\Theta$,

$$\sup_{\Lambda\in\mathcal{A}(\Theta)} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta = \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta} W_\theta(r) z_\theta.$$

(for a proof see Lemma A.3 in the appendix or Aubin and Ekeland, 1984, Theorem 6, Section 2, Chapter 6) □

Remark 3.1 *Note that the proof uses only what is known as a lower bound in the LDP.*

Remark 3.2 *Now we are in a position to explain why we consider condition* (U) *to be important in the definition of the dominated LDP. Assume that* $\{\mathcal{E}_n\,, n\geq 1\}$ *is a dominated family with dominating measures* P_n *such that, for deviability* V_Θ *on* R_+^Θ*, we have that* $\mathcal{L}(Z_{n,\Theta}|P_n)\stackrel{l.d.}{\to}V_\Theta$*. The proof of Theorem 3.1 with* $\mathbf{V}_\Theta$ *replaced by* V_Θ *and* $\mathbf{V}_\Lambda$ *replaced by* $V_\Theta\circ\pi_\Lambda^{-1}$ *(which would not use condition* (U)*) would still give the right-hand side of (3.2) as a lower bound. However these lower bounds can generally be different for different sequences of dominating measures. The role of condition* (U) *is to eliminate this possibility by making sure that equality (3.2) holds so that the lower bounds do not depend on the choice of dominating measures.*

In applications, as we will see, the assumption that the loss functions are level compact is normally met. However, in the appendix we give a variant of Theorem 3.1 for more general loss functions. This requires, as in the classical theory introducing generalised decisions, cf. Strasser, 1985.

We now turn our attention to the question of the tightness of the above lower bound and start with defining the concept of large deviation efficiency. Say that a sequence of decisions $\{\rho_n^*,\, n\geq 1\}$ is large deviation (LD) efficient, if for any other sequence of decisions $\{\rho_n\}$,

$$\overline{\lim_{n\to\infty}}\ \left(R_n(\rho_n^*)-R_n(\rho_n)\right)\leq 0.$$

Due to Theorem 3.1, to construct LD efficient decisions, one can apply an approach similar to the one used in the classical asymptotic decision theory. Indeed, by Theorem 3.1, if the $W_\theta\,,\theta\in\Theta$, are level compact, then, for any sequence of decisions $\{\rho_n,\, n\geq 1\}$,

$$\underline{\lim}_{n\to\infty}\ R_n(\rho_n)\geq R^*.$$

Now if a sequence $\{\rho_n^*,\, n\geq 1\}$ is such that $R_n(\rho_n^*)\to R^*$ as $n\to\infty$, it is obviously LD efficient.

Further, having in mind applications, we will be assuming that the sequence $\{\mathcal{E}_n\,, n\geq 1\}$ is dominated and conditions (Y) and (U) hold. Then, by Theorem 3.1, the asymptotic minimax risk can be written as

$$R^*=\sup_{y\in\mathcal{Y}}\inf_{r\in\mathcal{D}}\sup_{\theta\in\Theta}W_\theta(r)\mathfrak{z}_\theta(y)V(y). \tag{3.4}$$

Representation (3.4) prompts considering for each $y\in\mathcal{Y}$ the subproblem

$$(Q)\qquad Q^*(y)=\inf_{r\in\mathcal{D}}\sup_{\theta\in\Theta}W_\theta(r)\mathfrak{z}_\theta(y)\,.$$

Since the functions W_θ are level compact for each $\theta\in\Theta$, there exist $r^*(y)\in\mathcal{D}$, $y\in\mathcal{Y}$, which attain the inf. The value $r^*(y)$ can be viewed as "the best decision if the value of Y_n is y". Hence, provided the function $r^*(y):\,\mathcal{Y}\to\mathcal{D}$ is Borel, the decisions $r^*(Y_n)$ are natural candidates for LD efficient decisions. Unfortunately, we cannot prove this without requiring that $Q^*(y)$ be continuous (or upper semicontinuous) which is not usually the case in applications. The reason for the latter is that, as we have seen, $\mathfrak{z}_\theta(y)$ typically are not continuous as maps from $\mathcal{Y}$ into R_+. So we introduce continuous functions $Q_\delta(y)$ approximating $Q^*(y)$. Specifically, we define the subproblems

(Q_δ) $$Q_\delta(y) = \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta} W_\theta(r)\mathfrak{z}_{\theta,\delta}(y)\,,\; y\in\mathcal{Y}\,,$$

where $\mathfrak{z}_{\theta,\delta}(y)$ are, on the one hand, close to $\mathfrak{z}_\theta(y)$ and, on the other hand, such that $Q_\delta(y)$ is continuous. We achieve this through a stronger version of condition (Y) which we denote by $(\sup Y)$ and which requires, roughly, that (Y) hold uniformly in $\theta\in\Theta$. This way of handling the technical difficulties does not allow us, however, to get LD efficient decisions: as the next theorem shows, we are able to obtain only decisions whose asymptotic risk, in general, is arbitrarily close to the lower bound. Still we succeed in proving that the lower bound of Theorem 3.1 is tight and LD efficient decisions exist. We next state the condition. Recall that $Z_{n,\theta} = (dP_{n,\theta}/dP_n)^{1/n}$.

$(\sup Y)$ There exist statistics $Y_n : \Omega_n \to \mathcal{Y}$ with values in a metric space $\mathcal{Y}$ with Borel σ-field, functions $\mathfrak{z}_\theta : \mathcal{Y} \to R_+$, $\theta\in\Theta$, and $\mathfrak{z}_{\theta,\delta} : \mathcal{Y}\to R_+$, $\theta\in\Theta, \delta>0$, such that

$(Y.1)$ the sequence $\{\mathcal{L}(Y_n|P_n), n\ge 1\}$ LD converges to deviability $V(y)$, $y\in\mathcal{Y}$;

$(\sup Y.2)$ for uniform topology on R_+^Θ, the functions $\mathfrak{z}_{\Theta,\delta} = (\mathfrak{z}_{\theta,\delta}, \theta\in\Theta) : \mathcal{Y}\to R_+^\Theta$, $\delta>0$, are Borel and continuous V-a.e.;

$(\sup Y.3)$ $$\lim_{\delta\to 0}\ \overline{\lim_{n\to\infty}}\ \sup_{\theta\in\Theta} P_n^{1/n}\left(|Z_{n,\theta} - \mathfrak{z}_{\theta,\delta}(Y_n)| > \varepsilon\right) = 0,\ \varepsilon>0;$$

$(\sup Y.4)$ $$\lim_{\delta\to 0}\sup_{\theta\in\Theta}\ \sup_{y\in\Phi_V(a)} |\mathfrak{z}_{\theta,\delta}(y) - \mathfrak{z}_\theta(y)| = 0 \text{ for all } a>0.$$

In the next theorem, condition $(\sup Y)$ is used together with condition $(\sup U)$ which strengthens (U):

$(\sup U)$ $$\lim_{H\to\infty}\ \overline{\lim_{n\to\infty}}\ \sup_{\theta\in\Theta} E_n^{1/n} Z_{n,\theta}^n 1(Z_{n,\theta} > H) = 0.$$

Theorem 3.2 *Let a sequence of dominated experiments $\{\mathcal{E}_n,\ P_n,\ n\ge 1\}$ satisfy conditions $(\sup Y)$ and $(\sup U)$, and the function $W_\theta(r)$ be bounded in (θ, r) and level compact in r for each $\theta\in\Theta$. Assume that there exist Borel functions $r_\delta(y) : \mathcal{Y}\to\mathcal{D}$ such that the* inf *in (Q_δ) is attained at $r_\delta(y)$, and let $\rho_{n,\delta} = r_\delta(Y_n)$.*

Then

$$\lim_{\delta\to 0}\ \overline{\lim_{n\to\infty}}\ R_n(\rho_{n,\delta}) = \lim_{\delta\to 0}\ \underline{\lim_{n\to\infty}}\ R_n(\rho_{n,\delta}) = R^*,$$

so that

$$\lim_{n\to\infty}\ \inf_{\rho_n\in\mathcal{R}_n} R_n(\rho_n) = R^*.$$

In particular,

$$\lim_{n\to\infty} R_n(\rho_n^*) = R^*$$

for some sequence ρ_n^*.

Proof Since $(\sup Y)$ implies (Y), by Lemma 2.1, $\mathcal{L}(Z_{n,\Theta}|P_n) \stackrel{l.d.}{\to} V_\Theta = V \circ \mathfrak{z}_\Theta^{-1}$, so by Theorem 3.1, for each δ,

$$\varliminf_{n\to\infty} R_n(\rho_{n,\delta}) \geq R^*.$$

Proof of the first set of equalities would be over if

$$\varlimsup_{\delta\to 0} \varlimsup_{n\to\infty} R_n(\rho_{n,\delta}) \leq R^*. \tag{3.5}$$

Let C be an upper bound for W: $W_\theta(r) \leq C$. Since

$$R_n(\rho_{n,\delta}) = \sup_{\theta\in\Theta} E_{n,\theta}^{1/n} W_\theta^n(\rho_{n,\delta}) = \sup_{\theta\in\Theta} E_n^{1/n} W_\theta^n(\rho_{n,\delta}) Z_{n,\theta}^n,$$

we have that, for any $H > 0$,

$$R_n(\rho_{n,\delta}) \leq \sup_{\theta\in\Theta} E_n^{1/n} W_\theta^n(\rho_{n,\delta})(Z_{n,\theta}\wedge H)^n + C \sup_{\theta\in\Theta} E_n^{1/n} Z_{n,\theta}^n 1(Z_{n,\theta} > H).$$

The second term on the right tends to 0 as $n\to\infty$ and $H\to\infty$ by condition $(\sup U)$, so the required would follow by

$$\varlimsup_{\delta\to 0} \varlimsup_{n\to\infty} \sup_{\theta\in\Theta} E_n^{1/n} W_\theta^n(\rho_{n,\delta})(Z_{n,\theta}\wedge H)^n \leq R^*. \tag{3.6}$$

Since

$$\begin{aligned}|\sup_{\theta\in\Theta} E_n^{1/n} W_\theta^n(\rho_{n,\delta})(Z_{n,\theta}\wedge H)^n - \sup_{\theta\in\Theta} E_n^{1/n} W_\theta^n(\rho_{n,\delta})(\mathfrak{z}_{\theta,\delta}(Y_n)\wedge H)^n| \\ \leq C \sup_{\theta\in\Theta} E_n^{1/n}(|Z_{n,\theta} - \mathfrak{z}_{\theta,\delta}(Y_n)|\wedge H)^n,\end{aligned}$$

condition $(\sup Y.2)$ implies that

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} |\sup_{\theta\in\Theta} E_n^{1/n} W_\theta^n(\rho_{n,\delta})(Z_{n,\theta}\wedge H)^n - \sup_{\theta\in\Theta} E_n^{1/n} W_\theta^n(\rho_{n,\delta})(\mathfrak{z}_{\theta,\delta}(Y_n)\wedge H)^n| = 0. \tag{3.7}$$

Next, using the definitions of Q_δ and $\rho_{n,\delta}$, and the inequality $W_\theta(r)\leq C$, we get

$$\begin{aligned}&\sup_{\theta\in\Theta} E_n^{1/n} W_\theta^n(\rho_{n,\delta})(\mathfrak{z}_{\theta,\delta}(Y_n)\wedge H)^n \\ &\leq E_n^{1/n}\left(\sup_{\theta\in\Theta}(W_\theta^n(\rho_{n,\delta}(y))\mathfrak{z}_{\theta,\delta}(Y_n))\wedge CH\right)^n \\ &= E_n^{1/n}\left(Q_\delta(Y_n)\wedge CH\right)^n.\end{aligned} \tag{3.8}$$

The last two expectations in (3.8) are well defined since, by assumptions, $Q_\delta(y) = \sup_{\theta\in\Theta} W_\theta(r_\delta(y))\mathfrak{z}_{\theta,\delta}(y)$ is Borel.

By (Q_δ) and $(\sup Y.2)$, the function $Q_\delta(y)$ is V-a.e. continuous. Since $\mathcal{L}(Y_n|P_n) \stackrel{l.d.}{\to} V$, we get

$$\lim_{n\to\infty} E_n^{1/n}\left(Q_\delta(Y_n)\wedge CH\right)^n = \sup_{y\in\mathcal{Y}}(Q_\delta(y)\wedge CH)\,V(y). \tag{3.9}$$

By (Q), (Q_δ) and the inequality $W_\theta(r) \le C$, we have that

$$|\sup_{y\in\mathcal{Y}}(Q_\delta(y)\wedge CH)\,V(y) - \sup_{y\in\mathcal{Y}}(Q^*(y)\wedge CH)\,V(y)| \\ \le C\sup_{y\in\mathcal{Y}}\sup_{\theta\in\Theta}(|\mathfrak{z}_{\theta,\delta}(y)-\mathfrak{z}_\theta(y)|\wedge H)\,V(y),$$

and $(\sup Y.4)$ easily implies that the right-hand side tends to 0 as $\delta\to 0$. Thus

$$\lim_{\delta\to 0}\sup_{y\in\mathcal{Y}}(Q_\delta(y)\wedge CH)\,V(y) = \sup_{y\in\mathcal{Y}}(Q^*(y)\wedge CH)\,V(y) \\ \le \sup_{y\in\mathcal{Y}} Q^*(y)V(y) = R^*, \tag{3.10}$$

where for the last equality we used (3.4) and (Q). Putting together (3.7)–(3.10) proves (3.6) and hence (3.5).

The last claim of the theorem follows by (3.5) and a string of inequalities the first of which is Theorem 3.1

$$R^* \le \varliminf_{n\to\infty}\inf_{\rho_n} R_n(\rho_n) \le \varlimsup_{n\to\infty}\inf_{\rho_n} R_n(\rho_n) \le \varlimsup_{n\to\infty} R_n(\rho_{n,\delta}).$$

□

Remark 3.3 *Obviously, $r_\delta(y)$ chosen so that*

$$\sup_{\theta\in\Theta} W_\theta(r_\delta(y))\mathfrak{z}_{\theta,\delta}(y) \ge Q_\delta(y) - \epsilon_\delta,$$

where $\epsilon_\delta\to 0$ as $\delta\to 0$ would work too.

Remark 3.4 *If condition $(\sup Y)$ holds with $\mathfrak{z}_{\theta,\delta}(y)=\mathfrak{z}_\theta(y)$, then the $r_\delta(y)$ do not depend on δ and the decisions $\rho_n^* := \rho_{n,\delta}$ are LD efficient.*

Remark 3.5 *Assume that Θ is a topological space and denote by $C(\Theta,R)$ the subspace of R_+^Θ of continuous functions endowed with uniform topology. Then condition $(\sup Y.2)$ is implied by the following condition.*

$(\sup Y.2.1)$ *Θ is a compact metric space, the functions $\mathfrak{z}_{\theta,\delta}(y)\,,\delta>0$, are continuous in θ for each $y\in\mathcal{Y}$ and condition $(Y.2)$ holds.*

For a proof, note that under the assumptions Θ and $C(\Theta,R)$ are separable (see, e.g., Engelking, 1977, chapter 4) so that Borel σ-fields on $C(\Theta,R)$ for Tihonov and uniform topologies coincide.

Remark 3.6 *As with condition (Y), in applications, it is more convenient to deal with a logarithmic form of condition $(\sup Y)$. Namely, defining $\Xi_{n,\Theta}$ and $\Phi_I'(a)$ as in Remark 2.2, introduce condition $(\sup Y')$:*

$(\sup Y')$ *there exist statistics $Y_n : \Omega_n \to \mathcal{Y}$ with values in a metric space $\mathcal{Y}$ with Borel σ-field, functions $\zeta_\theta : \mathcal{Y}\to R$, $\theta\in\Theta$, and $\zeta_{\theta,\delta} : \mathcal{Y}\to R$, $\theta\in\Theta,\delta>0$, such that*

$(Y'.1)$ *the sequence* $\{\mathcal{L}(Y_n|P_n), n \geq 1\}$ *satisfies the LDP on* $\mathcal{Y}$ *with rate function* $I(y)$, $y \in \mathcal{Y}$;

$(\sup Y'.2)$ *for uniform topology on* R^Θ, *the functions* $\zeta_{\Theta,\delta} = (\zeta_{\theta,\delta}, \theta \in \Theta) : \mathcal{Y} \to R^\Theta$, $\delta > 0$, *are Borel and continuous at each point* y *such that* $I(y) < \infty$;

$(\sup Y'.3)$ $$\lim_{\delta \to 0} \overline{\lim_{n\to\infty}} \sup_{\theta\in\Theta} P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(Y_n)| > \varepsilon) = 0 \textit{ for all } \varepsilon > 0;$$

$(\sup Y'.4)$ $$\lim_{\delta\to 0} \sup_{\theta\in\Theta} \sup_{y\in\Phi'_I(a)} |\zeta_{\theta,\delta}(y) - \zeta_\theta(y)| = 0 \textit{ for all } a \geq 0.$$

Then $(\sup Y)$ *is implied by* $(\sup Y')$. *Similarly* $(\sup U)$ *follows from*

$(\sup U')$ $$\lim_{H\to\infty} \overline{\lim_{n\to\infty}} \sup_{\theta\in\Theta} E_n^{1/n} \exp(n\Xi_{n,\theta}) 1(\Xi_{n,\theta} > H) = 0.$$

Also $(\sup Y'.2)$ *is implied by*

$(\sup Y'.2.1)$ Θ *is a compact metric space, the functions* $\zeta_{\theta,\delta}(y)$, $\delta > 0$, *are continuous in* θ *for each* $y \in \mathcal{Y}$ *and condition* $(Y'.2)$ *holds.*

We will further be referring to the decisions $\rho_{n,\delta}$ as nearly LD efficient.

4 Asymptotic LD Risks and Efficient Decisions for Hypotheses Testing and Estimation Problems

This section specifies the above asymptotic minimax bound and (nearly) LD efficient decisions for typical statistical set-ups which are hypotheses testing and estimation with Bahadur-type criteria. We are considering indicator loss functions, i.e.,
$$W_\theta(r) = 1(r \notin A_\theta),\ r \in \mathcal{D}, \theta \in \Theta,$$
where A_θ are closed subsets of $\mathcal{D}$. Then the LD risk of decision ρ_n in the n-th experiment is
$$R_n(\rho_n) = \sup_{\theta\in\Theta} P_{n,\theta}^{1/n}(\rho_n \notin A_\theta).$$
For applications, it is handy to introduce the logarithmic risk
$$R'_n(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(\rho_n \notin A_\theta). \tag{4.1}$$
Accordingly, we consider the logarithm of the lower bound R^*:
$$R'^* = \sup_{\zeta_\Theta \in R^\Theta} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta: A_\theta \not\ni r} (\zeta_\theta - \mathbf{I}_\Theta(\zeta_\Theta)),$$
where $\mathbf{I}_\Theta(\zeta_\Theta) = -\log \mathbf{V}_\Theta(z_\Theta)$ for $z_\Theta = (\exp(\zeta_\theta), \theta \in \Theta)$, $\zeta_\Theta = (\zeta_\theta, \theta \in \Theta)$.

Theorem 3.1 then yields

Theorem 4.1 *Assume that the $A_\theta, \theta \in \Theta$, are compact. If the sequence $\{\mathcal{E}_n, n \geq 1\}$ obeys the LDP, then*

$$\varliminf_{n\to\infty} \inf_{\rho_n \in \mathcal{R}_n} R'_n(\rho_n) \geq R'^*.$$

Further, we will be assuming that the sequence $\{\mathcal{E}_n, n \geq 1\}$ is dominated, and conditions (Y') and (U') hold. According to Remark 2.2 and Theorem 3.1, we then have that

$$R'^* = \sup_{y\in\mathcal{Y}} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta\,:\,A_\theta \not\ni r} (\zeta_\theta(y) - I(y)). \tag{4.2}$$

Subproblems (Q) and (Q_δ) defined in Section 3 take the form

(Q') $$Q'^*(y) = \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta\,:\,A_\theta\not\ni r} \zeta_\theta(y)\,,\ y \in \mathcal{Y}\,,$$

and

(Q'_δ) $$Q'_\delta(y) = \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta\,:\,A_\theta\not\ni r} \zeta_{\theta,\delta}(y)\,,\ y \in \mathcal{Y}\,.$$

Obviously,

$$R'^* = \sup_{y\in\mathcal{Y}} (Q'^*(y) - I(y)).$$

Let the inf in (Q'_δ) be attained at points $r'_\delta(y)$ which is the case, e.g., if the $A_\theta\,, \theta \in \Theta$, are compact. We denote $\rho'_{n,\delta} = r'_\delta(Y_n)$.

Combining Theorem 4.1 and Theorem 3.2, and taking into account Theorem 3.1, Remarks 2.2 and 3.6, we obtain

Theorem 4.2 *Assume that $\{\mathcal{E}_n, P_n, n \geq 1\}$ is a dominated sequence of statistical experiments and the $A_\theta, \theta \in \Theta$, are compact.*

1. *If conditions (Y') and (U') hold, then*

$$\varliminf_{n\to\infty} \inf_{\rho_n\in\mathcal{R}_n} R'_n(\rho_n) \geq R'^*.$$

2. *Assume that the functions r'_δ, $\delta > 0$, mapping $\mathcal{Y}$ into $\mathcal{D}$, are Borel. If conditions $(\sup Y')$ and $(\sup U')$ hold, then*

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R'_n(\rho'_{n,\delta}) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R'_n(\rho'_{n,\delta}) = R'^*,$$

so that

$$\lim_{n\to\infty} \inf_{\rho_n\in\mathcal{R}_n} R'_n(\rho_n) = R'^*.$$

4.1 Hypotheses Testing

Let Θ_0 and Θ_1 be nonintersecting subsets of the parameter set Θ: $\Theta_0\subset\Theta, \Theta_1\subset\Theta, \Theta_0\cap\Theta_1=\emptyset$. We wish to test the hypothesis $H_0:\theta\in\Theta_0$ versus the alternative $H_1:\theta\in\Theta_1$.

The decision space $\mathcal{D}$ consists of two points: $\mathcal{D}=\{0,1\}$, we endow it with discrete topology, and, for any decision (test) ρ, we treat the event $\{\rho=0\}$ (respectively, $\{\rho=1\}$) as accepting (respectively, rejecting) the null hypothesis.

The corresponding loss function $W_\theta(r)$ is the indicator of the wrong choice:

$$W_\theta(r)=1(\theta\notin\Theta_r),\ r=0,1, \tag{4.3}$$

and the logarithmic risk $R'(\rho_n)$ of decision ρ_n from (4.1) takes the form

$$R_n^T(\rho_n)=\max\left\{\sup_{\theta\in\Theta_0}\frac{1}{n}\log P_{n,\theta}(\rho_n=1), \sup_{\theta\in\Theta_1}\frac{1}{n}\log P_{n,\theta}(\rho_n=0)\right\}. \tag{4.4}$$

Denoting the asymptotic minimax risk R'^* by T^*, we have by (4.2) that

$$T^*=\sup_{y\in\mathcal{Y}}\min\left\{\sup_{\theta\in\Theta_0}(\zeta_\theta(y)-I(y)), \sup_{\theta\in\Theta_1}(\zeta_\theta(y)-I(y))\right\}. \tag{4.5}$$

In applications, it is more convenient to use another representation for T^*, i.e.,

$$T^* \quad = \sup_{\theta\in\Theta_0,\,\theta'\in\Theta_1} S(\theta,\theta'), \tag{4.6}$$

where

$$S(\theta,\theta')=\sup_{y\in\mathcal{Y}}\min\left\{\zeta_\theta(y)-I(y),\zeta_{\theta'}(y)-I(y)\right\}. \tag{4.7}$$

Next, subproblem (Q'_δ) for this case is

$$T'_\delta(y)=\min_{r=0,1}\sup_{\theta\in\Theta_{1-r}}\zeta_{\theta,\delta}(y)\,,\ y\in\mathcal{Y}.$$

It has the solution

$$r_\delta^T(y)=1\left(\sup_{\theta\in\Theta_0}\zeta_{\theta,\delta}(y)<\sup_{\theta\in\Theta_1}\zeta_{\theta,\delta}(y)\right),$$

which leads to tests $\rho'_{n,\delta}$ of the form

$$\rho_{n,\delta}^T=1\left(\sup_{\theta\in\Theta_0}\zeta_{\theta,\delta}(Y_n)<\sup_{\theta\in\Theta_1}\zeta_{\theta,\delta}(Y_n)\right). \tag{4.8}$$

In the case of two simple hypotheses θ_0 and θ_1, the tests reduce to a regularised version of the Neyman-Pearson test:

$$\rho_{n,\delta}^T=1\left(\zeta_{\theta_0,\delta}(Y_n)<\zeta_{\theta_1,\delta}(Y_n)\right).$$

Thus Theorem 4.2 yields

Theorem 4.3 *Let* Θ_0 *and* Θ_1 *be nonintersecting subsets of* Θ. *If a sequence of dominated experiments* $\{\mathcal{E}_n,\, P_n,\, n \geq 1\}$ *satisfies conditions* (Y') *and* (U'), *then*

$$\varliminf_{n\to\infty} \inf_{\rho_n \in \mathcal{R}_n} R_n^T(\rho_n) \geq T^*.$$

If conditions $(\sup Y')$ *and* $(\sup U')$ *hold, then*

$$\lim_{n\to\infty} \inf_{\rho_n \in \mathcal{R}_n} R_n^T(\rho_n) = T^*,$$

and the tests $\rho_{n,\delta}^T$ *are nearly LD efficient:*

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = T^*.$$

4.2 Parameter Estimation

Let Θ be a subset of a normed space $\mathcal{B}$ with norm $\|\cdot\|$. We are interested in estimating parameter θ under the Bahadur-type loss function

$$W_\theta(r) = 1(\|r-\theta\| > c) \tag{4.9}$$

for given positive c. The logarithmic risk of estimator ρ_n is

$$R_n^E(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(\|\rho_n - \theta\| > c). \tag{4.10}$$

We assume that the decision space $\mathcal{D}$ is either a compact subset of $\mathcal{B}$ with induced topology, or a closed convex subset of $\mathcal{B}$ with weak topology (e.g., $\mathcal{D} = \mathcal{B}$); in the latter case, $\mathcal{B}$ is assumed to be a reflexive Banach space. The $W_\theta, \theta \in \Theta$, are then level compact on $\mathcal{D}$.

In this set-up, we denote the asymptotic minimax risk R'^* from (4.2) by E^*:

$$E^* = \sup_{y\in\mathcal{Y}} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta\,:\,\|r-\theta\|>c} (\zeta_\theta(y) - I(y)) \tag{4.11}$$

and the corresponding subproblem (Q'_δ) is

$$(E_\delta) \qquad E_\delta(y) = \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta\,:\,\|r-\theta\|>c} \zeta_{\theta,\delta}(y), \;\; y \in \mathcal{Y}.$$

We next describe solutions to (E_δ). Consider a real-valued function $f(\theta), \theta \in \Theta$, and let

$$A(h) = \{\theta\in\Theta:\ f(\theta) > h\},\ h \in R, \tag{4.12}$$

$$r(h) = \inf_{r\in\mathcal{D}} \sup_{\theta\in A(h)} \|r-\theta\|,\ h\in R, \tag{4.13}$$

$$h_c = \inf(h:\ r(h) \leq c).$$

We assume that $h_c < \infty$ (e.g., $f(\theta)$ is bounded). Note that, for both definitions of $\mathcal{D}$, $\inf_{r\in\mathcal{D}}$ in (4.13) is attained (for the case of weak topology, see, e.g., Baiocchi and Capelo, 1984, Theorem 2.2).

Lemma 4.1 *The set* $D_c = \{r \in \mathcal{D} : \sup_{\theta \in A(h_c)} \|r - \theta\| \le c\}$ *is nonempty and consists of all* $r_c \in \mathcal{D}$ *such that*

$$\sup_{\theta \in \Theta : ||r_c - \theta|| > c} f(\theta) = \inf_{r \in \mathcal{D}} \sup_{\theta \in \Theta : ||r - \theta|| > c} f(\theta),$$

where both sides are equal to h_c.

Proof It is not difficult to see that $r(h)$ is decreasing and right continuous. Hence $r(h_c) \le c$ and, since $\inf_{r \in \mathcal{D}} \sup_{\theta \in A(h_c)} ||r - \theta|| = r(h_c)$ and the inf is attained, the set D_c is nonempty.

Now let $r_c \in D_c$. By definition, $\|r_c - \theta\| \le c$ for all $\theta \in \Theta$ such that $f(\theta) > h_c$. Hence

$$\sup_{\theta \in \Theta : ||r_c - \theta|| > c} f(\theta) \le h_c. \tag{4.14}$$

On the other hand, if $h < h_c$, then $r(h) > c$ which implies that, for any $r \in \mathcal{D}$, $\sup_{\theta \in A(h)} \|r - \theta\| > c$ or, equivalently, there exists θ such that $f(\theta) > h$ and $\|r - \theta\| > c$, so that $\inf_{r \in \mathcal{D}} \sup_{\theta \in \Theta : ||r - \theta|| > c} f(\theta) \ge h$. Since h is arbitrarily close to h_c, we conclude that

$$\inf_{r \in \mathcal{D}} \sup_{\theta \in \Theta : ||r - \theta|| > c} f(\theta) \ge h_c,$$

which by (4.14) proves that r_c has the required property.

Finally, if $r \notin D_c$, then $\sup_{\theta \in A(h_c)} \|r - \theta\| > c$, i.e., there exists θ such that $\|r - \theta\| > c$ and $f(\theta) > h_c$ which yields the inequality $\sup_{\theta \in \Theta : ||r - \theta|| > c} f(\theta) > h_c$. $\square$

Remark 4.1 *Informally,* $r(h)$ *is the smallest radius of the balls which contain all the* θ *with* $f(\theta) > h$*, and* h_c *is the lowest level* h *for which there exists a ball of radius* c *with this property. The lemma states that* h_c *is the* inf *over all the balls of radii* c *of the largest values of* $f(\theta)$ *outside the balls.*

If we consider the case of one-dimensional parameter θ*, the construction in the lemma chooses the lowest level set of the function* f *which is contained in an interval of length* $2c$*, and the* r_c *are the centres of the intervals. Motivated by this interpretation, this type of estimators could be called* interval-median.

For given f, let $r_c(f)$ denote an element of the set D_c in the lemma and, taking $f(\theta) = \zeta_{\theta,\delta}(y)$, let $r^E_{\delta,c}(y) = r_c(\zeta_{\Theta,\delta}(y))$. We assume that the functions $r^E_{\delta,c}(y) : \mathcal{Y} \to \mathcal{D}$ are Borel. We can then define the estimators

$$\rho^E_{n,\delta} = r^E_{\delta,c}(Y_n). \tag{4.15}$$

A version of Theorem 4.2 for this case is

Theorem 4.4 *Assume that either* $\mathcal{B}$ *is a normed space and* $\mathcal{D}$ *is its compact subset with induced topology, or* $\mathcal{B}$ *is a reflexive Banach space and* $\mathcal{D}$ *is a closed convex subset of* $\mathcal{B}$ *with weak topology. Let* $\Theta \subset \mathcal{B}$.
If a sequence of dominated experiments $\{\mathcal{E}_n, P_n, n \ge 1\}$ *satisfies conditions* (Y') *and* (U')*, then*

$$\varliminf_{n \to \infty} \inf_{\rho_n \in \mathcal{R}_n} R^E_n(\rho_n) \ge E^*.$$

If conditions $(\sup Y')$ *and* $(\sup U')$ *hold, then*

$$\lim_{n\to\infty} \inf_{\rho_n\in\mathcal{R}_n} R_n^E(\rho_n) = E^*,$$

and the interval-median estimators $\rho_{n,\delta}^E = r_{\delta,c}^E(Y_n)$ *are nearly LD efficient,*

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}} R_n^E(\rho_{n,\delta}^E) = \lim_{\delta\to 0} \underline{\lim_{n\to\infty}} R_n^E(\rho_{n,\delta}^E) = E^*.$$

Remark 4.2 *If* $\mathcal{B}$ *is a separable reflexive Banach space, then the Borel* σ*–fields for strong and weak topologies coincide, hence the condition of the measurability of* $r_{\delta,c}^E$ *does not depend on a topology on* $\mathcal{B}$*.*

4.3 Estimation of Linear Functionals

Let Θ be a vector space and let $L(\cdot)$ be a linear functional on Θ. Consider the problem of estimating $L(\theta)$. We take $\mathcal{D} = R$, the real line. As above, we consider Bahadur-type criteria: the loss function is

$$W_\theta(r) = 1(|r - L(\theta)| > c),\ \theta\in\Theta, r\in R,$$

where $c > 0$ is fixed, and the risk of estimator ρ_n is given by

$$R_n^F(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}\left(|\rho_n - L(\theta)| > c\right). \tag{4.16}$$

The asymptotic minimax lower bound R'^* takes the form

$$F^* = \sup_{y\in\mathcal{Y}} \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta\,:\,|r-L(\theta)|>c} (\zeta_\theta(y) - I(y)), \tag{4.17}$$

and subproblem (Q'_δ) is

$$(F_\delta) \qquad F_\delta(y) = \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta\,:\,|r-L(\theta)|>c} \zeta_{\theta,\delta}(y),\ y\in\mathcal{Y}.$$

Corresponding solutions $\rho'_\delta(y)$ can be constructed along the same lines as for the parameter estimation problem above. Namely, fixing y and δ, denote $f(\theta) = \zeta_{\theta,\delta}(y)$ and let, for $h\in R$,

$$L\circ A(h) = \{L(\theta) : \theta\in A(h)\},$$

where $A(h)$ is from (4.12), be the image of $A(h)$ on the real line for the mapping L. Let $B(h)$ be the minimal closed interval in R containing $L\circ A(h)$. Set further, denoting by $d(B(h))$ the length of $B(h)$,

$$h_{c,L} \quad = \quad \inf\{h : d(B(h)) \le 2c\}.$$

Finally, consider intervals $B_{c,L}$ of length $2c$ which contain $B(h_{c,L})$ (note that $d(B(h_{c,L})) \le 2c$), and let $D_{c,L}$ be the set of the centres of all such intervals. The argument of Lemma 4.1 yields the following result.

Lemma 4.2 *The set $D_{c,L}$ is nonempty and consists of all $r_{c,L} \in \mathcal{D}$ such that*

$$\sup_{\theta\in\Theta:|r_{c,L}-L(\theta)|>c} f(\theta) = \inf_{r\in\mathcal{D}} \sup_{\theta\in\Theta:|r-L(\theta)|>c} f(\theta),$$

where both sides equal $h_{c,L}$.

To emphasise dependence on f, let us denote the elements of $D_{c,L}$ by $r_{c,L}(f)$. By the lemma, $r^F_{\delta,c}(y) = r_{c,L}(\zeta_{\Theta,\delta}(y))$ solves (F_δ). Assuming that the $r^F_{\delta,c}(y)$ are Borel functions from $\mathcal{Y}$ into R, we introduce the estimators $\rho^F_{n,\delta}$ of $L(\theta)$ by

$$\rho^F_{n,\delta} = r_{c,L}(\zeta_{\Theta,\delta}(Y_n)), \tag{4.18}$$

and call them also interval-median. Theorem 4.2 then yields

Theorem 4.5 *If a sequence of dominated experiments $\{\mathcal{E}_n,\, P_n,\, n\geq 1\}$ satisfies conditions (Y') and (U'), then*

$$\underline{\lim}_{n\to\infty} \inf_{\rho_n\in\mathcal{R}_n} R^F_n(\rho_n) \geq F^*.$$

If conditions $(\sup Y')$ and $(\sup U')$ hold, then

$$\lim_{n\to\infty} \inf_{\rho_n\in\mathcal{R}_n} R^F_n(\rho_n) = F^*,$$

and the interval-median estimators $\rho^F_{n,\delta} = r_{c,L}(\zeta_{\Theta,\delta}(Y_n))$ are nearly LD efficient,

$$\lim_{\delta\to 0} \overline{\lim}_{n\to\infty} R^F_n(\rho^F_{n,\delta}) = \lim_{\delta\to 0} \underline{\lim}_{n\to\infty} R^F_n(\rho^F_{n,\delta}) = F^*.$$

We conclude the section by giving a more explicit representation for F^*.

Lemma 4.3 *Under the above notation and conditions,*

$$F^* = \sup_{\theta,\theta'\,:\,|L(\theta-\theta')|>2c} S(\theta,\theta'),$$

where $S(\theta,\theta')$ is defined by (4.7):

$$S(\theta,\theta') = \sup_{y\in\mathcal{Y}} \min\left\{\zeta_\theta(y) - I(y), \zeta_{\theta'}(y) - I(y)\right\}.$$

Proof We fix $y \in \mathcal{Y}$ with $I(y) < \infty$, set $f(\theta) = \zeta_\theta(y)$ and define $h_{c,L}$ as above. We show that

$$h_{c,L} = \sup_{\theta,\theta'\,:\,|L(\theta-\theta')|>2c} \min\left\{f(\theta), f(\theta')\right\}.$$

By (4.17) and Lemma 4.2, this implies the claim.

Since $d(B(h)) \leq 2c$ for $h > h_{c,L}$, we have that, if $\theta,\theta' \in \Theta$ are such that $|L(\theta-\theta')| > 2c$, then $\min(f(\theta), f(\theta')) \leq h_{c,L}$. Conversely, if $h < h_{c,L}$, then $d(B(h)) > 2c$, hence there exist $\theta,\theta' \in \Theta$ such that $L(\theta-\theta') > 2c$ and $f(\theta) > h$, $f(\theta') > h$ which, by the arbitrariness of $h < h_{c,L}$, ends the proof. □

Remark 4.3 *The latter case of functional estimation includes the case of the estimation of one-dimensional parameter θ with $L(\theta) = \theta$, so the result of Lemma 4.3 applies to evaluating E^* too.*

5 Statistical Applications

In this section, we go back to the statistical models we introduced in Section 2 and apply to them the above general results. We first verify the LDP for the models by checking conditions (Y') and (U'). This is done under more general hypotheses than those of Section 2. After that, we present conditions which imply $(\sup Y')$ and $(\sup U')$. Next, considering certain hypotheses testing and estimation problems for the models, we calculate asymptotic minimax risks and indicate (nearly) LD efficient decisions.

Each of the subsections below uses its own notation. We mention it if certain symbols are re-used in different subsections for the same objects. For reader's convenience, we repeat the main points of the analysis we carried out for the models in Section 2 and recall the models themselves. Also we implicitly assume that the functions we choose as estimators are properly measurable.

5.1 Gaussian Observations

We observe a sample of n i.i.d. r.v. $\mathbf{X}_n = (X_{1,n}, \dots, X_{n,n})$ which are normally distributed with $\mathcal{N}(\theta, 1)$, $\theta \in \Theta \subset R$. For this model, $\Omega_n = R^n$ and $P_{n,\theta} = (\mathcal{N}(\theta,1))^n$, $\theta \in \Theta$. We take $P_{n,0}$ as dominating measure P_n. Then

$$\frac{1}{n} \log \frac{dP_{n,\theta}}{dP_n}(\mathbf{X}) = \frac{1}{n} \sum_{k=1}^{n} (\theta X_k - \frac{1}{2}\theta^2),\ \mathbf{X} = (X_1, \dots, X_n) \in R^n.$$

Thus it is natural to take

$$Y_n = \frac{1}{n} \sum_{k=1}^{n} X_{k,n},\ n \ge 1,$$

so that

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{dP_{n,\theta}}{dP_n}(\mathbf{X}_n) = \theta Y_n - \frac{1}{2}\theta^2.$$

Then $\{\mathcal{L}(Y_n|P_n), n \ge 1\}$ satisfies the LDP on R with the rate function $I^N(y) = y^2/2, y \in R$ (see, e.g., Freidlin and Wentzell, 1984). This checks condition $(Y'.1)$.

We next take

$$\zeta_\theta(y) = \zeta_{\theta,\delta}(y) = \theta y - \frac{1}{2}\theta^2. \tag{5.1}$$

Conditions $(Y'.2)$–$(Y'.4)$ are then obvious. Condition (U') follows by Chebyshev's inequality, since

$$E_n^{1/n} \exp(n\Xi_{n,\theta}) 1(\Xi_{n,\theta} > H) \le e^{-H} E_n^{1/n} \exp(2n\Xi_{n,\theta}) \to e^{-H} e^{3\theta^2}.$$

By Remark 2.2, the sequence $\{\mathcal{E}_n, n \ge 1\}$ obeys the LDP.

Let us assume further that Θ is bounded. It is then readily seen that conditions $(\sup Y')$ and $(\sup U')$ are met. We turn now to hypotheses testing and estimation problems and begin with calculating, for $\theta, \theta' \in \Theta$, the value $S(\theta, \theta')$ from (4.7).

Lemma 5.1 *For any* $\theta, \theta' \in \Theta$,

$$S(\theta, \theta') := \sup_{y \in R} \min \left\{ \zeta_\theta(y) - I^N(y), \zeta_{\theta'}(y) - I^N(y) \right\} = -\frac{(\theta - \theta')^2}{8}.$$

Proof By (5.1) and the definition of I^N, $\zeta_\theta(y) - I(y) = -(y-\theta)^2/2$, so

$$S(\theta, \theta') = \sup_{y \in R} \min \left\{ -\frac{(y-\theta)^2}{2}, -\frac{(y-\theta')^2}{2} \right\} = -\frac{(\theta - \theta')^2}{8}.$$

□

5.1.1 Testing $\theta = 0$ versus $|\theta| \geq 2c$

Assume that Θ contains 0 as an internal point. We are testing the simple hypothesis $H_0 : \theta = 0$ versus the two-sided alternative $H_1 : |\theta| \geq 2c$ with some prescribed $2c > 0$ such that the interval $[-2c, 2c]$ is contained in Θ .

The corresponding logarithmic risk of test ρ_n is given by (see (4.4))

$$R_n^T(\rho_n) = \max \left\{ \frac{1}{n} \log P_{n,0}(\rho_n = 1), \frac{1}{n} \sup_{|\theta| \geq 2c} \log P_{n,\theta}(\rho_n = 0) \right\}.$$

Now, using (4.6) with $\Theta_0 = \{0\}$ and $\Theta_1 = \{\theta \in \Theta : |\theta| \geq 2c\}$, and Lemma 5.1, we readily get

$$T^* = \sup_{|\theta'| \geq 2c} S(0, \theta') = -\frac{c^2}{2}.$$

Next, by Theorem 4.3 and Remark 3.4 to Theorem 3.2, LD efficient tests ρ_n^T can be taken in the form

$$\begin{aligned}
\rho_n^T &= 1\left(\sup_{|\theta| \geq 2c} \zeta_\theta(Y_n) > \zeta_0(Y_n) \right) \\
&= 1\left(\sup_{|\theta| \geq 2c} (\theta Y_n - \frac{\theta^2}{2}) > 0 \right) \\
&= 1(|Y_n| > c).
\end{aligned}$$

Applying Theorem 4.2, we arrive at the following result.

Proposition 5.1 *Let* $[-2c, 2c] \subset \Theta$. *Then*

$$\underline{\lim}_{n \to \infty} \inf_{\rho_n} R_n^T(\rho_n) \geq -\frac{c^2}{2}.$$

If Θ *is bounded, then*

$$\lim_{n \to \infty} \inf_{\rho_n} R_n^T(\rho_n) = -\frac{c^2}{2},$$

and the above tests ρ_n^T *are LD efficient:*

$$\lim_{n \to \infty} R_n^T(\rho_n^T) = -\frac{c^2}{2}.$$

5.1.2 Parameter Estimation

Now we consider the problem of estimating parameter θ. Recall (see (4.10)) that, given $c > 0$, the risk of estimator ρ_n is

$$R_n^E(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta| > c).$$

The value E^* of the asymptotic minimax risk is given by Lemma 4.3 (see Remark 4.3),

$$E^* = \sup_{\theta,\theta'\in\Theta\,:\,|\theta-\theta'|>2c} S(\theta,\theta').$$

By Lemma 5.1, we have that $E^* = -c^2/2$ if Θ contains an interval of length $2c$. An application of Theorem 4.4 and Remark 3.4 yields the following result.

Proposition 5.2 *Let Θ contain an interval of length $2c$. Then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^E(\rho_n) \geq -\frac{c^2}{2}.$$

If Θ is bounded, then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^E(\rho_n) = -\frac{c^2}{2},$$

and the interval-median estimators $\rho_n^E = r_c(\zeta_\Theta(Y_n))$ (see Section 4.2) are LD efficient:

$$\lim_{n\to\infty} R_n^E(\rho_n^E) = -\frac{c^2}{2}.$$

Remark 5.1 *It is easy to see that the estimator $\rho_n^E = r_c(\zeta_n(Y_n))$ coincides with Y_n if $Y_n - c \in \Theta$ and $Y_n + c \in \Theta$. Direct calculations show that the estimators $\rho_n = Y_n$ are also LD efficient, i.e., $\lim_n R_n^E(\rho_n) = -c^2/2$. The latter estimator is of simpler structure and does not depend on c and Θ. But the ρ_n^E seem to perform better at points outside or close to the boundary of Θ. In particular, if $Y_n \notin \Theta$, Θ is convex and is a subset of $\mathcal{D}$, then $\rho_n \notin \Theta$ whereas $\rho_n^E \in \Theta$.*

5.2 An I.I.D. Sample

We observe an i.i.d. sample $\mathbf{X}_n = (X_{1,n}, \ldots, X_{n,n})$ from distribution $P_\theta, \theta \in \Theta$. We assume that the family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ is dominated by probability measure P, i.e., $P_\theta \ll P$, $\theta \in \Theta$. This model is described by the dominated experiments $\mathcal{E}_n = (\Omega_n, \mathcal{F}_n; P_{n,\theta}, \theta \in \Theta)$ with $\Omega_n = R^n$, $\mathcal{F}_n = \mathcal{B}(R^n)$, $P_{n,\theta} = P_\theta^n$, $\theta \in \Theta$, $P_n = P^n$.

Assume that the family $\mathcal{P}$ satisfies the following regularity conditions:

$(R.1)$ *the densities $dP_\theta/dP(x), \theta \in \Theta$, are continuous and positive functions of $x \in R$;*

$(R.2)$ *the analogue of Cramér's condition holds,*

$$\int_R \left(\frac{dP_\theta}{dP}(x)\right)^\gamma P(dx) < \infty\,,\ \theta \in \Theta\,, \qquad \textit{for all } \gamma \in R.$$

We have that

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{dP_{n,\theta}}{dP_n}(\mathbf{X}_n) = \sum_{k=1}^n \frac{1}{n}\log\frac{dP_\theta}{dP}(X_{k,n}) = \int_R \log\frac{dP_\theta}{dP}(x)F_n(dx),$$

where

$$F_n(x) = \frac{1}{n}\sum_{k=1}^n 1(X_{k,n} \le x), \qquad x \in R, \tag{5.2}$$

are empirical distribution functions.

We take the latter as statistics Y_n. Then $\mathcal{Y}$ is the space of cumulative distribution functions on R which we denote by $\mathcal{F}$ and endow with the topology of weak convergence of corresponding probability measures. By Sanov's theorem, Sanov, 1957, Deuschel and Stroock, 1989, 3.2.17, the sequence $\{\mathcal{L}(Y_n|P_n), n \ge 1\}$ satisfies the LDP with $I^S(F) = K(F,P)\,,\ F \in \mathcal{F}$, where $K(F,P)$ is the Kullback-Leibler information:

$$K(F,P) = \begin{cases} \int\limits_R \frac{dF}{dP}(x)\log\frac{dF}{dP}(x)\,P(dx), & \text{if } F \ll P, \\ \infty, & \text{otherwise.} \end{cases} \tag{5.3}$$

This checks condition $(Y'.1)$. The verification of the rest of condition (Y') is more intricate than in the previous example.

Denote for $\theta \in \Theta$, $x \in R$ and $\delta > 0$,

$$\begin{aligned} L_\theta(x) &= \log\frac{dP_\theta}{dP}(x), \\ L_{\theta,\delta}(x) &= L_\theta(x) \wedge \delta^{-1} \vee (-\delta^{-1}), \end{aligned}$$

and define

$$\zeta_{\theta,\delta}(F) = \int_R L_{\theta,\delta}(x)F(dx), \quad F \in \mathcal{F}.$$

By $(R.1)$, the functions $\zeta_{\theta,\delta}$ are continuous on $\mathcal{F}$, so $(Y'.2)$ holds.

We check $(Y'.3)$. Condition $(R.2)$ implies that, for all $\gamma > 0$,

$$\lim_{\delta\to 0}\int_R \left[\exp\left(\gamma\left|L_\theta(x) - L_{\theta,\delta}(x)\right|\right) - 1\right] P(dx) = 0. \tag{5.4}$$

Then, for $\gamma > 0, \varepsilon > 0$, with the use of Chebyshev's inequality,

$$\begin{aligned} P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon) &\le P_n^{1/n}\left(\int_R |L_\theta(x) - L_{\theta,\delta}(x)|\,F_n(dx) > \varepsilon\right) \\ &\le \exp(-\gamma\varepsilon)E_n^{1/n}\exp\left(n\gamma\int_R |L_\theta(x) - L_{\theta,\delta}(x)|\,F_n(dx)\right) \\ &= \exp(-\gamma\varepsilon)\int_R \exp\left(\gamma\left|L_\theta(x) - L_{\theta,\delta}(x)\right|\right) P(dx). \end{aligned}$$

By (5.4), it then follows that

$$\overline{\lim_{\delta\to 0}}\ \overline{\lim_{n\to\infty}}\ P_n^{1/n}(|\Xi_{n,\theta}-\zeta_{\theta,\delta}(F_n)|>\varepsilon)\le \exp(-\gamma\varepsilon).$$

Since γ is arbitrary, $(Y'.3)$ follows.

We next check $(Y'.4)$ with

$$\zeta_\theta(F)=\begin{cases}\displaystyle\int_R L_\theta(x)F(dx), & \text{if } I^S(F)<\infty,\\ 0, & \text{otherwise.}\end{cases}\tag{5.5}$$

To begin with, we show that the ζ_θ are well defined. Since the functions $x\log x-x+1$ and $\exp x-1$ are convex conjugates (Rockafellar, 1970), by Young's inequality (see, e.g., Krasnoselskii and Rutickii, 1961), for $F\ll P$,

$$\begin{aligned}\int_R\left|L_\theta(x)\frac{dF}{dP}(x)\right|P(dx)&\le\int_R\left[\exp\left(|L_\theta(x)|\right)-1\right]P(dx)\\&+\int_R\left(\frac{dF}{dP}(x)\log\frac{dF}{dP}(x)-\frac{dF}{dP}(x)+1\right)P(dx)\\&\le 1+\int_R\left(\frac{dP_\theta}{dP}(x)\right)^{-1}P(dx)+I^S(F).\end{aligned}$$

In view of $(R.2)$, this proves that the ζ_θ are well defined.

Now, for F with $I^S(F)<\infty$, we have, for $\gamma>0$, using Young's inequality again, that

$$\begin{aligned}\gamma|\zeta_{\theta,\delta}(F)-\zeta_\theta(F)|&\le\int_R\gamma\left|L_{\theta,\delta}(x)-L_\theta(x)\right|F(dx)\\&\le\int_R\left[\exp\left(\gamma\left|L_{\theta,\delta}(x)-L_\theta(x)\right|\right)-1\right]P(dx)\\&+\int_R\left(\frac{dF}{dP}(x)\log\frac{dF}{dP}(x)-\frac{dF}{dP}(x)+1\right)P(dx)\\&=\int_R\left[\exp\left(\gamma\left|L_{\theta,\delta}(x)-L_\theta(x)\right|\right)-1\right]P(dx)+I^S(F).\end{aligned}$$

Hence by (5.4),

$$\overline{\lim_{\delta\to 0}}\ \sup_{F\in\Phi'_{I^S}(a)}|\zeta_{\theta,\delta}(F)-\zeta_\theta(F)|\le\frac{a}{\gamma},$$

and taking $\gamma\to\infty$, we arrive at $(Y'.4)$. Lemma 2.1 then implies that the LDP holds for $\{\mathcal{L}\left(\Xi_{n,\Theta}|P_n\right),n\ge 1\}$.

It remains to check (U'). Using once again Chebyshev's inequality, we obtain, for $H>0$,

$$\begin{aligned}E_n^{1/n}\exp(n\Xi_{n,\theta})1(\Xi_{n,\theta}>H)&\le\exp(-H)E_n^{1/n}\exp(2n\Xi_{n,\theta})\\&=\exp(-H)\int_R\left(\frac{dP_\theta}{dP}(x)\right)^2P(dx)\end{aligned}$$

and the assertion follows by condition $(R.2)$.

Conditions (Y') and (U') have been checked and thus the LDP holds.

Remark 5.2 *It is possible to do without condition* $(R.1)$. *Then bounded continuous functions* $L_{\theta,\delta}=(L_{\theta,\delta}(x), x\in R), \delta>0, \theta\in\Theta$, *should be chosen so that (5.4) holds. The existence of such functions follows from* $(R.2)$.

To check $(\sup Y')$ and $(\sup U')$, we assume that stronger versions of conditions $(R.1)$ and $(R.2)$ hold:

$(\sup R.1)$ *the functions* $dP_\theta/dP(x)$ *are positive and equicontinuous at each* $x\in R$;

$(\sup R.2)$ $\sup_{\theta\in\Theta}\int_R\left(\frac{dP_\theta}{dP}(x)\right)^\gamma P(dx)<\infty$ *for all* $\gamma\in R$.

Defining ζ_θ, $\zeta_{\theta,\delta}$, L_θ and $L_{\theta,\delta}$ as above, we have, by $(\sup R.2)$, that for all $\gamma>0$

$$\lim_{\delta\to 0}\sup_{\theta\in\Theta}\int_R\left[\exp\left(\gamma\left|L_\theta(x)-L_{\theta,\delta}(x)\right|\right)-1\right]P(dx)=0.$$

Using this, conditions $(\sup Y'.3)$ and $(\sup Y'.4)$ are checked as conditions $(Y'.3)$ and $(Y'.4)$ above. Condition $(\sup U')$ is also checked analogously to condition (U'), with the use of $(\sup R.2)$. Condition $(Y'.1)$ has already been checked.

It remains to check $(\sup Y'.2)$. We show that the functions $(\zeta_{\theta,\delta}(F), \theta\in\Theta)$ are continuous in F for uniform topology on R_+^Θ which obviously implies $(\sup Y'.2)$. Since weak topology on $\mathcal{F}$ is metrisable, it is enough to check sequential continuity. Let $F^{(n)}$ weakly converge to F as $n\to\infty$. Then the definition of the $L_{\theta,\delta}$ and $(\sup R.1)$ imply that the $L_{\theta,\delta}(x), \theta\in\Theta$, for δ fixed, are uniformly bounded and equicontinuous at each $x\in R$, so that (see, e.g., Billingsley, 1968, Problem 8, §2)

$$\sup_{\theta\in\Theta}\left|\int_R L_{\theta,\delta}(x)F^{(n)}(dx)-\int_R L_{\theta,\delta}(x)F(dx)\right|\to 0$$

checking $(\sup Y'.2)$. Conditions $(\sup Y')$ and $(\sup U')$ have been checked.

Remark 5.3 *The condition of equicontinuity in* $(\sup R.1)$ *holds if* Θ *is a compact topological space and the functions* $dP_\theta/dP(x)$, $\theta\in\Theta$, *are continuous in* θ *for each* x *and continuous in* x *for each* θ.

We now proceed to considering concrete statistical problems for the model. For this we need the following result by Chernoff, 1952, see also Kullback, 1959.

Lemma 5.2 *Let* $\mathcal{F}$ *be the space of all probability measures on a Polish space* E *with Borel* σ*–field and let measures* $P,Q\in\mathcal{F}$ *be dominated by measure* μ *with densities* $p(x)$ *and* $q(x)$. *Then*

$$\inf_{F\in\mathcal{F}}\max\{K(F,P),K(F,Q)\}=C(P,Q)$$

where $K(F,P)$ *is the Kullback-Leibler information (5.3) and* $C(P,Q)$ *is Chernoff's function*

$$C(P,Q)=-\inf_{\gamma\in[0,1]}\log\int p^\gamma(x)\,q^{1-\gamma}(x)\,\mu(dx).$$

We next apply Lemma 5.2 to calculate the function $S(\theta,\theta')$ from (4.7) which appears in expressions for minimax risks in hypotheses testing and estimation problems, see (4.6) and Lemma 4.3.

Lemma 5.3 *For* $\theta,\theta'\in\Theta$,

$$S(\theta,\theta') := \sup_{F\in\mathcal{F}} \min\left\{\zeta_\theta(F)-I^S(F), \zeta_{\theta'}(F)-I^S(F)\right\} = -C(P_\theta,P_{\theta'}).$$

Proof Let $I^S(F)<\infty$. Then $F \ll P$, and, since the densities $dP_\theta/dP(x)$, $\theta\in\Theta$, are positive, we also have that $F\ll P_\theta$ and, P-a.e.,

$$\frac{dF}{dP}=\frac{dF}{dP_\theta}\frac{dP_\theta}{dP}.$$

Therefore, by the definitions of ζ_θ and I^S,

$$\begin{aligned}\zeta_\theta(F)-I^S(F) &= \int_R \log\frac{dP_\theta}{dP}(x)F(dx)-\int_R\log\frac{dF}{dP}F(dx)\\ &= -\int_R \log\frac{dF}{dP_\theta}F(dx) = -K(F,P_\theta)\end{aligned}$$

and the result follows by Lemma 5.2. □

As a consequence of Theorem 4.3 and Lemma 5.3, we obtain the following result for a hypotheses testing problem. Consider the tests from (4.8):

$$\rho^T_{n,\delta} = 1\left(\sup_{\theta\in\Theta_0}\zeta_{\theta,\delta}(F_n) < \sup_{\theta\in\Theta_1}\zeta_{\theta,\delta}(F_n)\right).$$

As above, the risk $R^T_n(\rho_n)$ of test ρ_n is defined by (4.4). By (4.5) and Lemma 5.3,

$$T^* = -\inf_{\theta\in\Theta_0,\,\theta'\in\Theta_1} C(P_\theta,P_{\theta'}),$$

so Theorem 4.3 yields

Proposition 5.3 *Let* Θ_1 *and* Θ_2 *be nonintersecting subsets of* Θ. *If conditions* $(R.1)$ *and* $(R.2)$ *hold, then*

$$\varliminf_{n\to\infty}\ \inf_{\rho_n} R^T_n(\rho_n) \geq -\inf_{\theta\in\Theta_0,\,\theta'\in\Theta_1} C(P_\theta,P_{\theta'}).$$

If conditions $(\sup R.1)$ *and* $(\sup R.2)$ *hold, then*

$$\lim_{n\to\infty}\ \inf_{\rho_n} R^T_n(\rho_n) = -\inf_{\theta\in\Theta_0,\,\theta'\in\Theta_1} C(P_\theta,P_{\theta'}),$$

and the tests $\rho^T_{n,\delta}$ *are nearly LD efficient, i.e.,*

$$\begin{aligned}\lim_{\delta\to0}\ \varlimsup_{n\to\infty}\ R^T_n(\rho^T_{n,\delta}) &= \lim_{\delta\to0}\ \varliminf_{n\to\infty}\ R^T_n(\rho^T_{n,\delta})\\ &= -\inf_{\theta\in\Theta_0,\,\theta'\in\Theta_1} C(P_\theta,P_{\theta'}).\end{aligned}$$

In a similar manner one can tackle estimation problems for parameter θ or linear functionals of θ.

5.3 "Signal + White Noise"

We observe the stochastic process $X_n = (X_n(t), t \in [0,1])$ obeying the stochastic differential equation

$$dX_n(t) = \theta(t)dt + \frac{1}{\sqrt{n}}\, dW(t),\ 0 \le t \le 1, \tag{5.6}$$

where $W = (W(t), t \in [0,1])$ is a standard Wiener process and $\theta(\cdot)$ is an unknown function which we assume to be continuous.
This model is described by the statistical experiments $\mathcal{E}_n = (\Omega_n, \mathcal{F}_n; P_{n,\theta},\ \theta \in \Theta)$, where $\Omega_n = C[0,1]$, the space of continuous functions on $[0,1]$, $\Theta \subset C[0,1]$ and $P_{n,\theta}$ is the distribution of X_n on $C[0,1]$ for given θ. We take $P_n = P_{n,0}$, where $P_{n,0}$ corresponds to the zero function $\theta(\cdot) \equiv 0$. Then $P_{n,\theta} \ll P_n$ and, moreover, by Girsanov's formula, P_n-a.s.,

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{dP_{n,\theta}}{dP_n}(X_n) = \int_0^1 \theta(t) dX_n(t) - \frac{1}{2}\int_0^1 \theta^2(t)dt. \tag{5.7}$$

So, to check condition (Y'), we take $Y_n = X_n$ and $\mathcal{Y} = C[0,1]$ with uniform metric.

Let $C_0[0,1]$ be the subset in $C[0,1]$ of functions $x(\cdot)$ which are absolutely continuous w.r.t. Lebesgue measure and $x(0) = 0$. Since the sequence $\{\mathcal{L}(X_n | P_n), n \ge 1\}$ satisfies the LDP on $C[0,1]$ with the rate function

$$I^W(x(\cdot)) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^1 (\dot{x}(t))^2\, dt, & \text{if } x(\cdot) \in C_0[0,1] \\ \infty, & \text{otherwise,} \end{cases} \tag{5.8}$$

where $x(\cdot) \in C[0,1]$ and $\dot{x}(t)$ denotes the derivative of $x(\cdot)$ at t (see, e.g., Freidlin and Wentzell, 1984), condition $(Y'.1)$ holds.

We next take

$$\zeta_{\theta,\delta}(x(\cdot)) = \int_0^1 \theta_\delta(t)\, dx(t) - \frac{1}{2}\int_0^1 \theta^2(t)\, dt,\ x(\cdot) \in C[0,1], \tag{5.9}$$

where

$$\theta_\delta(t) = \sum_{k=0}^{[1/\delta]} \theta(k\delta) 1(t \in [k\delta, (k+1)\delta)),\ t \in [0,1], \tag{5.10}$$

and the first integral on the right of (5.9) is understood as a finite sum.

By the continuity of $\theta(\cdot)$,

$$\lim_{\delta \to 0} \int_0^1 (\theta(t) - \theta_\delta(t))^2\, dt = 0. \tag{5.11}$$

The $\zeta_{\theta,\delta}$ are obviously continuous in $x(\cdot) \in C[0,1]$, so $(Y'.2)$ holds. Next, by (5.7) and (5.9), we have, for $\varepsilon > 0$ and $\gamma > 0$, in view of Chebyshev's inequality, that

$$\begin{aligned} P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(X_n)| > \varepsilon) &\le P_n^{1/n}\left(\left|\int_0^1 (\theta(t) - \theta_\delta(t)) \frac{1}{\sqrt{n}}\, dW(t)\right| > \varepsilon\right) \\ &\le 2e^{-\gamma\varepsilon} \exp\left(\frac{\gamma^2}{2}\int_0^1 (\theta(t) - \theta_\delta(t))^2\, dt\right), \end{aligned}$$

and by (5.11)

$$\overline{\lim_{\delta\to 0}}\ \overline{\lim_{n\to\infty}}\ P_n^{1/n}(|\Xi_{n,\theta}-\zeta_{\theta,\delta}(X_n)|>\varepsilon)\le 2\exp(-\gamma\varepsilon),$$

which proves $(Y'.3)$ by the arbitrariness of γ.

For condition $(Y'.4)$, we take

$$\zeta_\theta(x(\cdot))=\begin{cases}\displaystyle\int_0^1\theta(t)\dot{x}(t)\,dt-\frac{1}{2}\int_0^1\theta^2(t)\,dt, & \text{if } I^W(x(\cdot))<\infty,\\ 0, & \text{otherwise.}\end{cases}$$

The ζ_θ are well defined, since by the Cauchy-Schwarz inequality and (5.8), if $x(\cdot)$ is absolutely continuous,

$$\int_0^1|\theta(t)\dot{x}(t)|\,dt\le\left(\int_0^1\theta^2(t)\,dt\right)^{1/2}(2I^W(x(\cdot)))^{1/2}.$$

Moreover, if $I^W(x(\cdot))<\infty$, then

$$\begin{aligned}|\zeta_{\theta,\delta}(x(\cdot))-\zeta_\theta(x(\cdot))|&\le\int_0^1|\theta_\delta(t)-\theta(t)||\dot{x}(t)|\,dt\\ &\le\left(\int_0^1(\theta_\delta(t)-\theta(t))^2\,dt\right)^{1/2}\left(\int_0^1(\dot{x}(t))^2\,dt\right)^{1/2},\end{aligned}$$

so

$$\sup_{x(\cdot)\in\Phi'_{I^W}(a)}|\zeta_{\theta,\delta}(x(\cdot))-\zeta_\theta(x(\cdot))|\le(2a)^{1/2}\left(\int_0^1(\theta_\delta(t)-\theta(t))^2\,dt\right)^{1/2},$$

and the latter goes to 0 as $\delta\to 0$ by (5.11). Condition (Y') has been verified.

It remains to check (U'). Using the model equation (5.6), (5.7) and Chebyshev's inequality again, we have that

$$\begin{aligned}E_n^{1/n}\exp(n\Xi_{n,\theta})1(\Xi_{n,\theta}>H)&\le\exp(-H)E_n^{1/n}\exp(2n\Xi_{n,\theta})\\ &=\exp(-H)\exp\left(\int_0^1\theta^2(t)dt\right)\to 0\qquad\text{as } H\to\infty.\end{aligned}$$

Conditions (Y') and (U') have been checked.

Remark 5.4 *The condition of the continuity of the $\theta(\cdot)$ can be weakened to the condition*

$$\int_0^1\theta^2(t)\,dt<\infty.$$

Functions θ_δ should then be chosen as step functions for which (5.11) holds.

For conditions $(\sup Y')$ and $(\sup U')$, we assume that the $\theta(\cdot)$ belong to a compact Σ in $C[0,1]$, more specifically, $\Sigma=\Sigma_0(\beta,M)$ which is a subset of the Hölder class

$$\Sigma(\beta,M)=\{\theta(\cdot):\ |\theta(t)-\theta(s)|\le M|t-s|^\beta,\forall s,t\in[0,1]\},\tag{5.12}$$

for some $\beta \in (0,1]$ and $M > 0$, with the additional property that

$$\sup_{\theta(\cdot)\in\Sigma_0(\beta,M)} |\theta(0)| < \infty.$$

The compactness of $\Sigma_0(\beta,M)$ in $C[0,1]$ is an obvious consequence of Arzelà–Ascoli's theorem. The conditions on $\Sigma_0(\beta,M)$ easily imply that

$$\sup_{\theta(\cdot)\in\Sigma_0(\beta,M)} \int_0^1 \theta^2(t)\,dt < \infty \tag{5.13}$$

and

$$\lim_{\delta\to 0} \sup_{\theta(\cdot)\in\Sigma_0(\beta,M)} \int_0^1 (\theta(t)-\theta_\delta(t))^2\,dt = 0. \tag{5.14}$$

Now conditions $(\sup Y'.3)$ and $(\sup Y'.4)$ are checked as conditions $(Y'.3)$ and $(Y'.4)$, respectively, with the use of (5.14) in place of (5.11). Condition $(\sup Y'.2)$ is checked as for the i.i.d. sample model since the $\zeta_{\theta,\delta}(x(\cdot)), \theta \in \Sigma_0(\beta,M)$, are equicontinuous at each x which easily follows from the compactness of $\Sigma_0(\beta,M)$. Finally, condition $(\sup U')$ follows in analogy with condition (U') with the use of (5.13). This completes the verification of conditions $(\sup Y')$ and $(\sup U')$.

We now calculate the function $S(\theta,\theta')$ from (4.7) for the model.

Lemma 5.4 *For any $\theta,\theta' \in C[0,1]$,*

$$\begin{aligned} S(\theta,\theta') &:= \sup_{x(\cdot)\in C[0,1]} \min\{\zeta_\theta(x(\cdot)) - I^W(x(\cdot)), \zeta_{\theta'}(x(\cdot)) - I^W(x(\cdot))\} \\ &= -\frac{1}{8}\int_0^1 [\theta(t)-\theta'(t)]^2 dt. \end{aligned}$$

Proof Since by the definitions of I^W and ζ_θ, for $x(\cdot)$ with $I^W(x(\cdot)) < \infty$,

$$\zeta_\theta(x(\cdot)) - I^W(x(\cdot)) = -\frac{1}{2}\int_0^1 (\dot{x}(t)-\theta(t))^2\,dt,$$

we get, using the inequality $\max(a^2,b^2) \ge (a-b)^2/4$,

$$\begin{aligned} S(\theta,\theta') \ge -\max\left\{\frac{1}{2}\int_0^1 [\dot{x}(t)-\theta(t)]^2 dt, \frac{1}{2}\int_0^1 [\dot{x}(t)-\theta'(t)]^2 dt\right\} \\ \ge -\frac{1}{8}\int_0^1 [\theta(t)-\theta'(t)]^2 dt. \end{aligned}$$

But for $x(\cdot)$ with $\dot{x}(t) = [\theta(t)+\theta'(t)]/2$, we have that

$$\frac{1}{2}\int_0^1 [\dot{x}(t)-\theta(t)]^2 dt = \frac{1}{2}\int_0^1 [\dot{x}(t)-\theta'(t)]^2 dt = -\frac{1}{8}\int_0^1 [\theta(t)-\theta'(t)]^2 dt$$

and the required follows. $\square$

Now we apply these formulae and the general results from Section 4 to two statistical problems concerning the value of the function $\theta(\cdot)$ at an internal point t_0 of $[0,1]$.

5.3.1 Testing $\theta(t_0) = 0$ versus $|\theta(t_0)| \geq 2c$

Given $c > 0$, denote $\Theta_0 = \{\theta \in \Theta : \theta(t_0) = 0\}$, $\Theta_1 = \{\theta \in \Theta : |\theta(t_0)| \geq 2c\}$ and define the risk $R_n^T(\rho_n)$ of test ρ_n by (4.4).

Proposition 5.4 *Let c, M and t_0 be such that $[t_0 - t^*, t_0 + t^*] \subseteq [0,1]$, where $t^* = (c/M)^{1/\beta}$.*
If $\Theta = \Sigma(\beta, M)$, then

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) \geq -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)} \left(\frac{c}{M}\right)^{1/\beta}.$$

If $\Theta = \Sigma_0(\beta, M)$, then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)} \left(\frac{c}{M}\right)^{1/\beta},$$

and the tests $\rho_{n,\delta}^T$ from (4.8) are nearly LD efficient, i.e.,

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)} \left(\frac{c}{M}\right)^{1/\beta}.$$

Proof We need only calculate T^* from (4.6). Denote

$$\theta^*(t) = [c - M|t - t_0|^\beta]^+. \tag{5.15}$$

If now $\theta \in \Theta_0, \theta' \in \Theta_1$, then the inequality $|\theta(t_0) - \theta'(t_0)| \geq 2c$ and the Hölder constraints (5.12) imply that $|\theta(t) - \theta'(t)| \geq 2[c - M|t-t_0|^\beta]^+ = 2\theta^*(t)$ and hence

$$\int_0^1 (\theta(t) - \theta'(t))^2 dt \geq \int_0^1 4(\theta^*(t))^2 dt.$$

This yields by Lemma 5.15

$$S(\theta,\theta') \leq -\frac{1}{8} 4 \int_0^1 (\theta^*(t))^2 \, dt = -\int_0^{t^*} (c - Mt^\beta)^2 dt$$
$$= -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)} \left(\frac{c}{M}\right)^{1/\beta}.$$

On the other hand, evidently, $c - \theta^* \in \Theta_0$, $c + \theta^* \in \Theta_1$ and $S(c - \theta^*, c + \theta^*) = -\frac{1}{2}\int_0^1 (\theta^*(t))^2 \, dt$. This proves the assertion. $\square$

5.3.2 Estimating $\theta(t_0)$

Treating the value $\theta(t_0)$ as a linear functional of $\theta(\cdot)$, we define the risk of estimator ρ_n of $\theta(t_0)$ by

$$R_n^F(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta(t_0)| > c).$$

Proposition 5.5 *Let c, M and t_0 be such that $[t_0 - t^*, t_0 + t^*] \subseteq [0,1]$. If $\Theta = \Sigma(\beta, M)$, then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) \geq -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)} \left(\frac{c}{M}\right)^{1/\beta}.$$

If $\Theta = \Sigma_0(\beta, M)$, then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)} \left(\frac{c}{M}\right)^{1/\beta},$$

and the interval-median estimators $\rho_{n,\delta}^F$ from (4.18) are nearly LD efficient, i.e.,

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R_n^T(\rho_{n,\delta}^F) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R_n^T(\rho_{n,\delta}^F) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)} \left(\frac{c}{M}\right)^{1/\beta}.$$

Proof By Theorem 4.5 and Lemma 4.3,

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) \geq F^* = \sup_{\theta,\theta' : |\theta(t_0) - \theta'(t_0)| > 2c} S(\theta, \theta').$$

Repeating the above calculation for the testing problem, we obtain, for $\theta^*(t) = [c - M|t-t_0|^\beta]^+$,

$$F^* = S(\theta^*, -\theta^*) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)} \left(\frac{c}{M}\right)^{1/\beta}.$$

$\square$

Remark 5.5 *The latter problem has been studied by Korostelev, 1993 who suggests different upper estimators, namely, the kernel estimators*

$$\hat{\rho}_n = \int K(t_0 - t) dX_n(t)$$

with the kernel $K(t) = (\beta+1)/(2c\beta)\,(M/c)^{1/\beta}\,[c - M|t-t_0|]^+$. These estimators have proved to be asymptotically efficient in the sense that $R_n^T(\hat{\rho}_n) \to F^*$ as $n \to \infty$.

5.4 Gaussian Regression

We are considering the regression model

$$X_{k,n} = \theta(t_{k,n}) + \xi_{k,n}, \qquad t_{k,n} = \frac{k}{n}, \quad k = 1, \ldots, n, \tag{5.16}$$

where errors $\xi_{k,n}$ are i.i.d. standard normal and $\theta(\cdot)$ is an unknown function which again is assumed to be continuous.

In this model, $\Omega_n = R^n$, $\Theta \subset C[0,1]$ and $P_{n,\theta}$ is the distribution of $\mathbf{X}_n = (X_{1,n}, \ldots, X_{n,n})$ for $\theta(\cdot)$. As above, we take $P_n = P_{n,0}$. Then

$$\begin{aligned}\Xi_{n,\theta} &= \frac{1}{n}\log\frac{dP_{n,\theta}}{dP_n}(X_n)\\ &= \frac{1}{n}\sum_{k=1}^{n}\theta(t_{k,n})X_{k,n} - \frac{1}{n}\sum_{k=1}^{n}\theta^2(t_{k,n})\\ &= \int_0^1 \theta(t)\,dX_n(t) - \frac{1}{n}\sum_{k=1}^{n}\theta^2\left(t_{k,n}\right),\end{aligned} \tag{5.17}$$

where

$$X_n(t) = \frac{1}{n}\sum_{k=1}^{[nt]} X_{k,n}, \qquad 0 \le t \le 1.$$

This prompts taking the process $X_n = (X_n(t)\,,\ t\in[0,1])$ as statistic Y_n in condition (Y'). Space $\mathcal{Y}$ is the space of right continuous with left-hand limits functions on $[0,1]$ with uniform metric (for the measurability of X_n, see Billingsley, 1968, §8).

Since the $X_{k,n}$ are $\mathcal{N}(0,1)$-distributed under P_n, the sequence $\{\mathcal{L}\left(X_n|P_n\right), n \ge 1\}$ satisfies the LDP with I^W from (5.8) (see, e.g., Puhalskii, 1994a). This checks condition $(Y'.1)$.

Next, we define $\zeta_{\theta,\delta}(x(\cdot))$ and $\theta_\delta(t)$ as in Subsection 5.3, i.e.,

$$\zeta_{\theta,\delta}(x(\cdot)) = \int_0^1 \theta_\delta(t)dx(t) - \frac{1}{2}\int_0^1 \theta^2(t)dt,\ x(\cdot)\in\mathcal{Y}, \tag{5.18}$$

$$\theta_\delta(t) = \sum_{k=0}^{[1/\delta]}\theta(k\delta)1(t\in[k\delta,(k+1)\delta))\,,\ t\in[0,1].$$

Note that the $\zeta_{\theta,\delta}$ are measurable w.r.t. the Borel σ-field on $\mathcal{Y}$ and continuous V^W-a.e. since they are continuous at continuous functions and $V^W(x(\cdot)) = 0$ if $x(\cdot)$ is not absolutely continuous. This checks condition $(Y'.2)$.

Now, by (5.17) and (5.18),

$$\begin{aligned}P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(X_n)| > \varepsilon) \le 1\left(\left|\int_0^1 \theta^2(t)\,dt - \frac{1}{n}\sum_{k=1}^{n}\theta^2(k/n)\right| > \varepsilon/2\right)\\ + P_n^{1/n}\left(\left|\int_0^1(\theta(t)-\theta_\delta(t))\,dX_n(t)\right| > \varepsilon/2\right).\end{aligned}$$

The first term on the right is zero for all n large enough by the continuity of $\theta(\cdot)$. The second is not greater than

$$\begin{aligned}&e^{-\gamma\varepsilon/2}E_n^{1/n}\exp\left(n\gamma\left|\int_0^1(\theta(t)-\theta_\delta(t))\,dX_n(t)\right|\right)\\ &\le 2e^{-\gamma\varepsilon/2}\exp\left(\frac{\gamma^2}{2n}\sum_{k=1}^{n}\left(\theta(k/n)) - \theta_\delta(k/n)\right)^2\right).\end{aligned}$$

By the continuity of $\theta(\cdot)$ and since the $\theta_\delta(\cdot)$ are step functions,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\left(\theta(k/n)-\theta_\delta(k/n)\right)^2=\frac{1}{2}\int_0^1\left(\theta(t)-\theta_\delta(t)\right)^2dt,$$

and the latter goes to 0 as $\delta\to 0$ by the continuity of $\theta(\cdot)$. Since γ is arbitrary, condition $(Y'.3)$ is checked.

Conditions $(Y'.4)$ and (U') are checked as for the "signal+ white noise" model (with the same choice of ζ_θ).

Remark 5.6 *As in the preceding model, instead of the continuity of* $\theta(\cdot)$, *we could require that it be square integrable on* $[0,1]$.

To get nearly LD efficient decisions, we assume that the $\theta(\cdot)$ belong to the class $\Sigma_0(\beta,M)$ defined above. Conditions $(\sup Y'.3)$, $(\sup Y'.4)$ and $(\sup U')$ again are checked as for the "signal + white noise" model if we take into account that

$$\lim_{n\to\infty}\sup_{\theta(\cdot)\in\Sigma_0(\beta,M)}\int_0^1\left(\theta([nt]/n)-\theta(t)\right)^2dt=0.$$

Condition $(\sup Y'.2)$ is obvious.

Since here we have the same functions $I^W(x)$ and $\zeta_\theta(x)$ as for the "signal + white noise" model, the statistical problems of Subsection 5.3 have the same solution.

5.5 Non-Gaussian Regression

We consider the same regression model (5.16) but now assume that the i.i.d. errors $\xi_{k,n}$ have distribution P with positive probability density function $p(x)$ w.r.t. Lebesgue measure on the real line. The unknown regression function $\theta(\cdot)$ is again assumed to be continuous, so $\Theta\subset C[0,1]$.

Next, we assume that the density $p(x)$ obeys the following condition, cf. conditions $(R.1)$ and $(R.2)$ for an i.i.d. sample:

(P) *the density* $p(x)$ *is positive, continuous and the function*

$$H_\gamma(s)=\int_R p^\gamma(x)p^{1-\gamma}(x-s)\,dx$$

is bounded and continuous in $s\in R$ *for all* $\gamma\in R$.

Again, for a regression function $\theta(\cdot)$, we denote by $P_{n,\theta}$ the distribution of $X_n=(X_{1,n},\ldots,X_{n,n})$. We have, with $P_n=P_{n,0}$,

$$\Xi_{n,\theta}=\frac{1}{n}\log\frac{dP_{n,\theta}}{dP_n}(X_n)=\frac{1}{n}\sum_{k=1}^{n}\log\frac{p(X_{k,n}-\theta(k/n))}{p(X_{k,n})}.$$

This representation suggests, as in the case of an i.i.d. sample, taking for Y_n the empirical process $F_n = F_n(x,t)$, $x \in R$, $t \in [0,1]$, defined by

$$F_n(x,t) = \frac{1}{n}\sum_{k=1}^{[nt]} 1(X_{k,n} \le x). \tag{5.19}$$

Then

$$\Xi_{n,\theta} = \int_0^1 \int_R \log \frac{p(x-\theta(t))}{p(x)} F_n(dx,dt). \tag{5.20}$$

We define $\mathcal{Y}$ as the space of cumulative distribution functions $F = F(x,t)$, $x \in R$, $t \in [0,1]$, on $R \times [0,1]$ with weak topology. Let $\mathcal{Y}_0$ be the subset of $\mathcal{Y}$ of absolutely continuous w.r.t. Lebesgue measure on $R \times [0,1]$ functions $F(x,t)$ with densities $p_t(x)$ satisfying the condition $\int_R p_t(x)\,dx = 1$, $\forall t \in [0,1]$. By Puhalskii, 1995c, the sequence $\{\mathcal{L}(F_n|P_n), n \ge 1\}$ obeys the LDP on $\mathcal{Y}$ with the rate function $I^{SK}(F)$ given by

$$I^{SK}(F) = \begin{cases} \displaystyle\int_0^1 \int_R \log \frac{p_t(x)}{p(x)}\, p_t(x)\,dx\,dt, & \text{if } F \in \mathcal{Y}_0, \\ \infty, & \text{otherwise.} \end{cases}$$

This checks $(Y'.1)$.

To define $\zeta_{\theta,\delta}(F)$, introduce the functions

$$\begin{aligned} L_\theta(x,t) &= \log \frac{p(x-\theta(t))}{p(x)}, \\ L_{\theta,\delta}(x,t) &= L_\theta(x,t) \vee (-\delta^{-1}) \wedge \delta^{-1},\ x \in R,\ t \in [0,1]. \end{aligned}$$

The functions $L_{\theta,\delta}$ are bounded, continuous and are such that

$$\lim_{\delta\to 0} \int_0^1 \int_R [\exp(\gamma\,|L_\theta(x,t) - L_{\theta,\delta}(x,t)|) - 1]\, p(x)\,dx\,dt = 0, \quad \gamma > 0. \tag{5.21}$$

We set

$$\zeta_{\theta,\delta}(F) = \int_0^1 \int_R L_{\theta,\delta}(x,t)\, F(dx,dt). \tag{5.22}$$

Then condition $(Y'.2)$ holds by the definition of topology on $\mathcal{Y}$ and the choice of the $L_{\theta,\delta}$.

For condition $(Y'.3)$, write, for $\gamma > 0$, using Chebyshev's inequality, and (5.19), (5.20) and (5.22),

$$\begin{aligned} &\frac{1}{n}\log P_n(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon) \\ &\le \frac{1}{n}\log P_n\left(\int_0^1 \int_R |L_\theta(x,t) - L_{\theta,\delta}(x,t)|\ F_n(dx,dt) > \varepsilon\right) \\ &\le -\gamma\varepsilon + \frac{1}{n}\sum_{k=1}^n \log \int_R \exp(\gamma\,|L_\theta(x,k/n) - L_{\theta,\delta}(x,k/n)|)\, p(x)\,dx. \end{aligned}$$

By condition (P), the continuity of $\theta(\cdot)$, and the boundedness and continuity of $L_{\theta,\delta}$, the second term on the right converges, as $n \to \infty$, to

$$\int_0^1 \log \int_R \exp\left(\gamma \left|L_\theta(x,t) - L_{\theta,\delta}(x,t)\right|\right) p(x)\, dx\, dt.$$

Since (5.21) implies, by Jensen's inequality, that the latter goes to 0 as $n \to \infty$, we conclude that

$$\overline{\lim_{\delta\to 0}}\ \overline{\lim_{n\to\infty}}\ \frac{1}{n} \log P_n(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon) \le -\gamma\varepsilon,$$

which proves $(Y'.3)$ since γ is arbitrary.

For condition $(Y'.4)$, we take

$$\zeta_\theta(F) = \begin{cases} \displaystyle\int_0^1 \int_R L_\theta(x,t)\, F(dx,dt), & \text{if } I^{SK}(F) < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

The ζ_θ are well defined since by Young's inequality, if $F(x,t) = \int_0^t \int_{-\infty}^x p_t(x) dx\, dt$, then

$$\begin{aligned} &\int_0^1 \int_R |L_\theta(x,t)| \frac{p_t(x)}{p(x)} p(x)\, dx\, dt \\ &\le \int_0^1 \int_R \left[\exp\left(|L_\theta(x,t)|\right) - 1\right] p(x)\, dx\, dt \\ &+ \int_0^1 \int_R \left(\frac{p_t(x)}{p(x)} \log \frac{p_t(x)}{p(x)} - \frac{p_t(x)}{p(x)} + 1 \right) p(x)\, dx\, dt \\ &\le 1 + \int_0^1 \int_R p^2(x) (p(x - \theta(t)))^{-1}\, dx\, dt + I^{SK}(F), \end{aligned}$$

which is finite, if $I^{SK}(F) < \infty$, by condition (P).

Next, with the use of Young's inequality, we have, for $\gamma > 0$, if $I^{SK}(F) < \infty$, that

$$\begin{aligned} \gamma|\zeta_{\theta,\delta}(F) - \zeta_\theta(F)| &\le \int_0^1 \int_R \gamma \left|L_{\theta,\delta}(x,t) - L_\theta(x,t)\right| F(dx,dt) \\ &\le \int_0^1 \int_R \left[\exp\left(\gamma \left|L_{\theta,\delta}(x,t) - L_\theta(x,t)\right|\right) - 1\right] p(x)\, dx\, dt + I^{SK}(F), \end{aligned}$$

so by (5.21),

$$\overline{\lim_{\delta\to 0}} \sup_{F \in \Phi'_{I^{SK}}(a)} |\zeta_{\theta,\delta}(F) - \zeta_\theta(F)| \le \frac{a}{\gamma},$$

which proves $(Y'.4)$ since γ is arbitrary.

Condition (U') is checked as in the case of an i.i.d. sample.

Remark 5.7 *We could weaken condition (P) to the condition*

$$\int_0^1 \int_R [p(x)]^\gamma [p(x-\theta(t))]^{1-\gamma}\,dx\,dt < \infty \ \ \textit{for all } \gamma \in R.$$

Then $L_{\theta,\delta}$ should be chosen so that (5.21) holds.

We now check conditions $(\sup Y')$ and $(\sup U')$. For this, we assume that the $\theta(\cdot)$ are again from the set $\Sigma_0(\beta, M)$ defined in Subsection 5.3. Then (5.21) can be strengthened to

$$\lim_{\delta\to 0} \sup_{\theta(\cdot)\in\Sigma_0(\beta,M)} \int_0^1 \int_R \left(\exp\left(\gamma\,|L_\theta(x,t) - L_{\theta,\delta}(x,t)|\right) - 1\right) p(x)\,dx\,dt = 0,\ \gamma > 0, \tag{5.23}$$

which allows us to check $(\sup Y'.3)$, $(\sup Y'.4)$ and $(\sup U')$ as $(Y'.3)$, $(Y'.4)$ and (U'), respectively. Condition $(\sup Y'.2)$ follows from the fact that the $L_{\theta,\delta}(x,t), \theta \in \Sigma_0(\beta, M)$, are equicontinuous at each (x,t), so the $(\zeta_{\theta,\delta}, \theta \in \Theta) : \mathcal{Y} \to R_+^\Theta$ are continuous for uniform topology on R_+^Θ.

We now calculate the function $S(\theta,\theta')$, $\theta,\theta' \in \Theta$ from (4.7). This is done with the use of a generalisation of Chernoff's result in Lemma 5.2 which we state and prove next. Let $(E,\mathcal{E})$ be a Polish space with Borel σ-field and let $\mathcal{P}(E)$ be the space of probability measures on $(E,\mathcal{E})$. As above, for $F, P \in \mathcal{P}(E)$, we denote by $K(F,P)$ the Kullback-Leibler information:

$$K(F,P) = \begin{cases} \displaystyle\int_E \log\frac{dF}{dP}(x) F(dx), & \text{if } F \ll P, \\ \infty, & \text{otherwise.} \end{cases}$$

Recall that $K(F,P)$, for P fixed, is convex and is a rate function in F for weak topology on $\mathcal{P}(E)$, Deuschel and Stroock, 1989, 3.2.17.

Obviously we can consider $E \times [0,1]$ with product topology in place of E. In this case, for a Borel measure ν on $[0,1]$, denote by $\mathcal{P}_\nu(E\times[0,1])$ the subset of $\mathcal{P}(E\times[0,1])$ of measures F such that $F(E\times[0,t]) = \nu([0,t])$, $t \in [0,1]$.

Our version of Chernoff's result is the following lemma.

Lemma 5.5 *Let E be a Polish space. Let probability measures $P, Q \in \mathcal{P}(E\times[0,1])$ be dominated by the product measure $\mu\times\nu$, where μ and ν are Borel measures on E and $[0,1]$ respectively.*

Then

$$\inf_{F\in\mathcal{P}_\nu(E\times[0,1])} \max\{K(F,P), K(F,Q)\}$$
$$= -\inf_{\gamma\in[0,1]} \int_0^1 \log\left[\int_E p_t^\gamma(x)\, q_t^{1-\gamma}(x)\,\mu(dx)\right] \nu(dt),$$

where $p_t(x)$ and $q_t(x)$ are the respective densities of P and Q relative to $\mu\times\nu$.

Proof Obviously,

$$\max\{K(F,P),K(F,Q)\} = \sup_{\gamma\in[0,1]} (\gamma K(F,P)+(1-\gamma)K(F,Q)). \tag{5.24}$$

Let $\mathcal{P}(E\times[0,1])$ be endowed with weak topology. Since $K(F,P)$ is convex and is a rate function in F, we deduce that the function $\gamma K(F,P)+(1-\gamma)K(F,Q)$, $\gamma\in[0,1], F\in\mathcal{P}_\nu(E\times[0,1])$, meets the conditions of a minimax theorem (see, e.g., Aubin and Ekeland, 1984, Theorem 7, Section 2, Chapter 6). Hence

$$\begin{aligned}&\inf_{F\in\mathcal{P}_\nu(E\times[0,1])}\sup_{\gamma\in[0,1]}(\gamma K(F,P)+(1-\gamma)K(F,Q))\\&=\sup_{\gamma\in[0,1]}\inf_{F\in\mathcal{P}_\nu(E\times[0,1])}(\gamma K(F,P)+(1-\gamma)K(F,Q)).\end{aligned} \tag{5.25}$$

The latter inf can equivalently be taken over F which are dominated by P and Q, and hence by $\mu\times\nu$. Denote by $f_t(x)$ the density of $F\ll\mu\times\nu$. Since, by the definition of $\mathcal{P}_\nu(E\times[0,1])$,

$$F(E\times[0,t])=\int_0^t\int_E f_t(x)\,\mu(dx)\,\nu(dt)=\nu([0,t]),\ t\in[0,1],$$

we have that

$$\int_E f_t(x)\,\mu(dx)=1 \qquad \nu\text{–a.e.} \tag{5.26}$$

Next, by the definition of the Kullback-Leibler information,

$$\begin{aligned}&\gamma K(F,P)+(1-\gamma)K(F,Q)\\&=\int_0^1\int_E\log\frac{f_t(x)}{p_t^\gamma(x)q_t^{1-\gamma}(x)}f_t(x)\,\mu(dx)\,\nu(dt),\end{aligned} \tag{5.27}$$

where $0/0=0$. Since the function $x\log x\,, x\geq 0$, is convex, an application of Jensen's unequality and (5.26) gives that ν–a.e. in $t\in[0,1]$

$$\int_E\log\frac{f_t(x)}{p_t^\gamma(x)q_t^{1-\gamma}(x)}f_t(x)\,\mu(dx)\geq-\log\int_E p_t^\gamma(x)q_t^{1-\gamma}(x)\,\mu(dx)\,.$$

On the other hand, taking

$$f_t(x)=p_t^\gamma(x)q_t^{1-\gamma}(x)\left(\int_E p_t^\gamma(x)q_t^{1-\gamma}(x)\,\mu(dx)\right)^{-1} \tag{5.28}$$

we obviously get equality above. Since the measure F with the density defined by (5.28) belongs to $\mathcal{P}_\nu(E\times[0,1])$, we obtain by (5.27) that

$$\begin{aligned}&\inf_{F\in\mathcal{P}_\nu(E\times[0,1])}[\gamma K(F,P)+(1-\gamma)K(F,Q)]\\&=-\int_0^1\log\left[\int_E p_t^\gamma(x)q_t^{1-\gamma}(x)\,\mu(dx)\right]\,\nu(dt)\,,\end{aligned}$$

which, by (5.24) and (5.25), concludes the proof. □

Remark 5.8 *Obviously, the* inf *in the statement can equivalently be taken over* $F \in \mathcal{P}_\nu(E \times [0,1])$ *such that* $K(F,P) < \infty$, $K(F,Q) < \infty$.

Remark 5.9 *Chernoff's result follows if* ν *is a Dirac measure.*

Now we apply Lemma 5.5 for evaluating the function $S(\theta, \theta')$.

Lemma 5.6 *For any* $\theta, \theta' \in \Theta$,

$$S(\theta, \theta') = \inf_{\gamma \in [0,1]} \int_0^1 \log H_\gamma(\theta'(t) - \theta(t))\, dt.$$

Proof We have, for $F \in \mathcal{Y}_0$ with $I^{SK}(F) < \infty$, that

$$\zeta_\theta(F) - I^{SK}(F) = -K(F, \overline{P}_\theta)\,,$$

where $\overline{P}_\theta(dx, dt) = p(x - \theta(t))\, dx\, dt$, and the claim follows by Lemma 5.5 and Remark 5.8 with $E = R\,, \mu(dx) = dx\,, \nu(dt) = dt\,, P = \overline{P}_\theta\,, Q = \overline{P}_{\theta'}$. □

The latter result enables us to calculate the value of minimax risks for various statistical problems. To compare with the Gaussian case, let us consider the same statistical problems dealing with the value of $\theta(t_0)$ for given t_0.

5.5.1 Testing $\theta(t_0) = 0$ versus $|\theta(t_0)| \geq 2c$

Given $c > 0$, denote $\Theta_0 = \{\theta \in \Theta : \theta(t_0) = 0\}$, $\Theta_1 = \{\theta \in \Theta : |\theta(t_0)| \geq 2c\}$ and define the risk $R_n^T(\rho_n)$ of test ρ_n by (4.4).

Proposition 5.6 *Let* c, M *and* t_0 *be such that* $[t_0 - t^*, t_0 + t^*] \subseteq [0,1]$ *where* $t^* = (c/M)^{1/\beta}$. *Let the measure* P *satisfy condition* (P) *and let the function* $H_\gamma(s)$ *monotonously increase in* $s \geq 0$.
If $\Theta = \Sigma(\beta, M)$, *then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) \geq \inf_{\gamma \in [0,1]} 2 \int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\, dt.$$

If $\Theta = \Sigma_0(\beta, M)$, *then*

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) = \inf_{\gamma \in [0,1]} 2 \int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\, dt,$$

and the tests $\rho_{n,\delta}^T$ *from (4.8) are nearly LD efficient, i.e.,*

$$\begin{aligned}\lim_{\delta\to 0} \varlimsup_{n\to\infty} \; R_n^T(\rho_{n,\delta}^T) &= \lim_{\delta\to 0} \varliminf_{n\to\infty} \; R_n^T(\rho_{n,\delta}^T) \\ &= \inf_{\gamma\in[0,1]} 2 \int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\, dt.\end{aligned}$$

Proof The result follows from Theorem 4.3 and we only need to evaluate T^* from (4.6). A straightforward calculation using Lemma 5.6 and the monotonicity of $H_\gamma(s)$ shows that

$$T^* := \sup_{\theta\in\Theta_0,\,\theta'\in\Theta_1} S(\theta,\theta') = \inf_{\gamma\in[0,1]} 2\int_0^1 \log H_\gamma(2\theta^*(t))\,dt,$$

where $\theta^*(t) = [c - M|t-t_0|^\beta]^+$. This obviously yields the claim by Lemma 5.6. □

5.5.2 Estimating $\theta(t_0)$

For the problem of the estimation of $\theta(t_0)$, the risk of estimator ρ_n is defined by

$$R_n^F(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta(t_0)| > c).$$

Proposition 5.7 *Let the conditions of Proposition 5.6 hold.*
If $\Theta = \Sigma(\beta, M)$, then

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) \ge \inf_{\gamma\in[0,1]} 2\int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\,dt.$$

If $\Theta = \Sigma_0(\beta, M)$, then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) = \inf_{\gamma\in[0,1]} 2\int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\,dt,$$

and the interval-median estimators $\rho_{n,\delta}^F$ from (4.18) are nearly LD efficient, i.e.,

$$\begin{aligned}\lim_{\delta\to 0} \varlimsup_{n\to\infty} R_n^F(\rho_{n,\delta}^F) &= \lim_{\delta\to 0} \varliminf_{n\to\infty} R_n^F(\rho_{n,\delta}^F) \\ &= \inf_{\gamma\in[0,1]} 2\int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\,dt.\end{aligned}$$

Proof Again it suffices to calculate the value of the asymptotic minimax risk given by Lemma 4.3,

$$F^* = \sup_{\theta,\theta'\in\Theta\,:\,|\theta(t_0)-\theta'(t_0)|>2c} S(\theta,\theta'),$$

which is done as for the "signal + white noise" model. □

Remark 5.10 *The latter problem of estimating $\theta(t_0)$ has been considered by Korostelev and Spokoiny, 1995 under the assumption that $\log p(x)$ is concave upward, and by Korostelev and Leonov, 1995 who study the double asymptotics as $n\to\infty$ and then $c\to 0$.*

5.6 The Change-Point Model

Let us observe a sample $X_n = (X_{1,n}, \ldots, X_{n,n})$ of real valued r.v., where, for some $k_n \geq 1$, the observations $X_{1,n}, \ldots, X_{k_n,n}$ are i.i.d. with distribution P_0 and the observations $X_{k_n+1,n}, \ldots, X_{n,n}$ are i.i.d. with distribution P_1. We are assuming that P_0 and P_1 are known and k_n is unknown. Also assume that $k_n = [n\theta]$, where $\theta \in \Theta = [0,1]$. Here $\Omega_n = R^n$, and $P_{n,\theta}$ denotes the distribution of X_n for given θ.

Let probability measure P dominate P_0 and P_1, and let

$$p_0(x) = \frac{dP_0}{dP}(x), \; p_1(x) = \frac{dP_1}{dP}(x), \quad x \in R,$$

be respective densities. We assume that $p_0(x)$ and $p_1(x)$ are positive and continuous, and

$$\int_R p_0^\gamma(x)\, P(dx) < \infty, \; \int_R p_1^\gamma(x)\, P(dx) < \infty \quad \text{for all } \gamma \in R. \tag{5.29}$$

Denoting $P_n = P^n$, we have

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{dP_{n,\theta}}{dP_n}(X_n) = \frac{1}{n} \sum_{i=1}^{[n\theta]} \log p_0(X_{i,n}) + \frac{1}{n} \sum_{i=[n\theta]+1}^{n} \log p_1(X_{i,n}),$$

so that defining an empirical process again by

$$F_n(x,t) = \frac{1}{n} \sum_{i=1}^{[nt]} 1(X_{i,n} \leq x), \; x \in R, t \in [0,1],$$

we obtain the representation

$$\Xi_{n,\theta} = \int_0^\theta \int_R \log p_0(x)\, F_n(dx,dt) + \int_\theta^1 \int_R \log p_1(x)\, F_n(dx,dt).$$

We define statistics Y_n and space $\mathcal{Y}$ as for the preceding model. Let $\mathcal{Y}_P$ consist of those $F \in \mathcal{Y}$ which are absolutely continuous relative to the measure $P(dx) \times dt$ and admit density $p_t(x)$ such that $\int_R p_t(x) P(dx) = 1\,, t \geq 0\,.$ As above, condition $(Y'.1)$ holds with

$$I_P^{SK}(F) = \begin{cases} \displaystyle\int_0^1 \int_R p_t(x) \log p_t(x) P(dx)\, dt, & \text{if } F \in \mathcal{Y}_P, \\ \infty, & \text{otherwise.} \end{cases}$$

We next take, for $F(\cdot,\cdot) \in \mathcal{Y}$,

$$\zeta_{\theta,\delta}(F) = \int_0^\theta \int_R L_{\delta,0}(x)\, F(dx,dt) + \int_\theta^1 \int_R L_{\delta,1}(x)\, F(dx,dt)\,,$$

where

$$L_{\delta,i}(x) = \log p_i(x) \wedge \delta^{-1} \vee (-\delta^{-1})\,, \; i = 0, 1.$$

The $L_{\delta,i}$ are bounded, continuous and

$$\lim_{\delta\to 0}\int_R \left[\exp\left(\gamma|\log p_i(x) - L_{\delta,i}(x)|\right) - 1\right] P(dx) = 0,\ i = 0, 1,\ \gamma > 0. \tag{5.30}$$

The $\zeta_{\theta,\delta}$ are easily seen to be Borel; also the continuity of $F(x,t)$ in t, if $V_P^{SK}(F) > 0$, implies that the $\zeta_{\theta,\delta}$ are V_P^{SK}-a.e. continuous. This checks $(Y'.2)$.

For $(Y'.3)$, write, by Chebyshev's inequality,

$$\begin{aligned} &P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon)\\ &\le P_n^{1/n}\left(\int_0^1\int_R |\log p_0(x) - L_{\delta,0}(x)|F_n(dx,dt) > \frac{\varepsilon}{2}\right)\\ &+ P_n^{1/n}\left(\int_0^1\int_R |\log p_1(x) - L_{\delta,1}(x)|F_n(dx,dt) > \frac{\varepsilon}{2}\right)\\ &\le \exp(-\gamma\varepsilon/2)\left[E_n^{[n\theta]/n}\exp(\gamma|\log p_0(X_{1,n}) - L_{\delta,0}(X_{1,n})|)\right.\\ &\left.+ E_n^{1-[n\theta]/n}\exp(\gamma|\log p_1(X_{1,n}) - L_{\delta,1}(X_{1,n})|)\right], \end{aligned}$$

so

$$\begin{aligned} &\overline{\lim_{n\to\infty}}\ P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon)\\ &\le \exp(-\gamma\varepsilon/2)\left[\left(\int_R \exp(\gamma|\log p_0(x) - L_{\delta,0}(x)|)P(dx)\right)^{\theta}\right.\\ &\left.+ \left(\int_R \exp(\gamma|\log p_1(x) - L_{\delta,1}(x)|)P(dx)\right)^{1-\theta}\right], \end{aligned}$$

and, by (5.30), this goes to $2\exp(-\gamma\varepsilon/2)$ as $\delta \to 0$. Condition $(Y'.3)$ is checked.

To check $(Y'.4)$, we take

$$\zeta_\theta(F) = \begin{cases} \displaystyle\int_0^\theta\int_R \log p_0(x)F(dx,dt) + \int_\theta^1\int_R \log p_1(x)F(dx,dt), & \text{if } I_P^{SK}(F) < \infty,\\ 0, & \text{otherwise.}\end{cases}$$

The fact that the ζ_θ are well defined and $(Y'.4)$ holds, is proved as for the non-Gaussian regression model. Condition (U') also is easily checked.

Remark 5.11 *The conditions on $p_0(x)$ and $p_1(x)$ can be weakened to the requirement that only (5.29) hold. One should then choose $L_{\delta,i}$ bounded, continuous and satisfying (5.30).*

Next, the argument used for (Y') and (U') checks also conditions $(\sup Y')$ and $(\sup U')$ (for $(\sup Y'.2)$ use condition $(\sup Y'.2.1)$ in Remark 3.6).

The next step is evaluating $S(\theta,\theta')$ for $\theta,\theta' \in [0,1]$.

Lemma 5.7 *For any $\theta,\theta' \in [0,1]$,*

$$S(\theta,\theta') = -|\theta - \theta'|\,C(P_0,P_1).$$

Proof In a manner similar to the case of non-Gaussian regression, we have, for any $F \in \mathcal{Y}_P, I_P^{SK}(F) < \infty$ with $F(dx,dt) = p_t(x)\,P(dx)\,dt$, that

$$\zeta_\theta(F) - I_P^{SK}(F) = -\int_0^\theta \int_R \log \frac{p_t(x)}{p_0(x)} p_t(x)\,P(dx)\,dt$$
$$-\int_\theta^1 \int_R \log \frac{p_t(x)}{p_1(x)} p_t(x)\,P(dx)\,dt = -K(F, \overline{P}_\theta),$$

where $\overline{P}_\theta(dx,dt) = (p_0(x)1(t \le \theta) + p_1(x)1(t > \theta))\,P(dx)\,dt$ and the claim follows by Lemma 5.5 and Remark 5.8 with $E = R\,, \mu(dx) = P(dx)\,,\ \nu(dt) = dt\,,\ P = \overline{P}_\theta\,,$ $Q = \overline{P}_{\theta'}\,.$ $\square$

We apply this result and the general theorems from Section 4 to the problem of the estimation of parameter θ. The risk of estimator ρ_n is defined in a standard way,

$$R_n^F(\rho_n) = \sup_{\theta\in[0,1]} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta| > c). \tag{5.31}$$

Proposition 5.8 *For any* $c < 1/2$*,*

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) = -2c\,C(P_0, P_1).$$

If $\rho_{n,\delta}^F$ *are the interval-median estimators from (4.18), then*

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}}\ R_n^F(\rho_{n,\delta}^F) = \lim_{\delta\to 0} \underline{\lim_{n\to\infty}}\ R_n^F(\rho_{n,\delta}^F) = -2c\,C(P_0, P_1).$$

Proof We apply Theorem 4.5. One only needs to calculate the value of the minimax risk F^*. Using Lemmas 4.3 and 5.7, we obtain

$$F^* = \sup_{\theta,\theta' : |\theta-\theta'| > 2c} S(\theta, \theta') = -2c\,C(P_0, P_1).$$

$\square$

Remark 5.12 *The same result has been obtained by Korostelev, 1995 who uses another kind of upper estimator. The construction is based on considering the concave hull of a sample path of the likelihood process. By Lemma 4.2 this estimator is a particular case of the interval-median estimators* $\rho_{n,\delta}^F$.

5.7 Regression with Random Design

We consider the model

$$X_{k,n} = \theta(t_{k,n}) + \xi_{k,n}, \qquad k = 1, \ldots, n, \tag{5.32}$$

where real-valued errors $\xi_{k,n}$ are i.i.d. with common distribution P having density $p(x)$ which obeys condition (P) of Subsection 5.5, and design points $t_{k,n}$ also are real-valued i.i.d. with common distribution Π and are independent of the $\xi_{k,n}$. We impose a standard condition on the design measure Π.

(II) *The measure* Π *is compactly supported and has positive density w.r.t. Lebesgue measure on the support.*

We denote the support by D. The unknown regression function $\theta(\cdot)$ is assumed to be continuous. In this model, $P_{n,\theta}$ is the joint distribution of $X_n = (X_{1,n}, \ldots, X_{n,n})$ and $t_n = (t_{1,n}, \ldots, t_{n,n})$ for θ.

Let us take for Y_n the joint empirical distribution function F_n of X_n and t_n:

$$F_n(A, B) = \frac{1}{n} \sum_{k=1}^{n} 1(X_{k,n} \in A, t_{k,n} \in B), \tag{5.33}$$

for Borel sets $A \subset R$, $B \subset D$. Space $\mathcal{Y}$ is the space of probability distributions on $R \times D$ submitted with weak topology. Set also $P_n = P_{n,0} = (P \times \Pi)^n$.

With these definitions,

$$\begin{aligned}
\Xi_{n,\theta} &= \frac{1}{n} \log \frac{dP_{n,\theta}}{dP_n}(X_n, t_n) \\
&= \frac{1}{n} \sum_{k=1}^{n} \log \frac{p(X_{k,n} - \theta(t_{k,n}))}{p(X_{k,n})} \\
&= \int_R \int_R \log \frac{p(x - \theta(t))}{p(x)} F_n(dx, dt).
\end{aligned}$$

Let $\mathcal{Y}_1$ be the subset of the set $\mathcal{Y}$ of two dimensional distribution functions on R^2 which are absolutely continuous w.r.t. Lebesgue measure on R^2 and have support in $R \times D$.

Under P_n, the random pairs $(X_{k,n}, t_{k,n})$ are i.i.d. with distribution $P \times \Pi$, and hence, by Sanov's theorem, the LDP holds for F_n with rate function $I^{SS}(F)$ defined by

$$I^{SS}(F) = \begin{cases} \displaystyle\int_R \int_D \log \frac{p(x,t)}{p(x)\pi(t)}\, p(x,t)\, dx\, dt, & \text{if } F \in \mathcal{Y}_1, \\ \infty, & \text{otherwise.} \end{cases}$$

Here $F(dx, dt) = p(x,t) dx dt$. This checks $(Y'.1)$.

Set next, for $F \in \mathcal{Y}$,

$$\begin{aligned}
\zeta_\theta(F) &= \begin{cases} \displaystyle\int_R \int_D \log \frac{p(x - \theta(t))}{p(x)} F(dx, dt), & \text{if } I^{SS}(F) < \infty, \\ 0, & \text{otherwise,} \end{cases} \\
\zeta_{\theta,\delta}(F) &= \int_R \int_D \left[\log \frac{p(x - \theta(t))}{p(x)}\right] \wedge \delta^{-1} \vee (-\delta^{-1})\, F(dx, dt).
\end{aligned}$$

With this notation, conditions (Y') and (U') are checked as for the non-Gaussian regression. This proves the LDP for the model.

For conditions $(\sup Y')$ and $(\sup U')$, we again assume, for the unknown regression function $\theta(\cdot)$, that $\theta \in \Sigma_0(\beta, M)$ with the same subset $\Sigma_0(\beta, M)$ of $\Sigma(\beta, M)$ as above. The conditions are then checked as for the non-Gaussian regression.

Now we are calculating the function $S(\theta, \theta')$ from (4.7). Recall that the function $H_\gamma(s)$ is defined in condition (P).

Lemma 5.8 *Under conditions* (P) *and* (Π),

$$S(\theta,\theta')=\inf_{\gamma\in[0,1]}\log\int_D H_\gamma(\theta'(t)-\theta(t))\pi(t)\,dt.$$

Proof Given $F\in\mathcal{Y}_1$ with $I^{SS}(F)<\infty$, we easily get

$$\zeta_\theta(F)-I^{SS}(F)=-K(F,\overline{P}),$$

where $\overline{P}_\theta(dx,dt)=p(x-\theta(t))\pi(t)\,dx\,dt$, and the claim follows by Lemma 5.2 with $E=R\times D$, $\mu(dx,dt)=dx\,dt$, $P=\overline{P}_\theta$, $Q=\overline{P}_{\theta'}$. $\square$

Now we again consider the same two statistical problems as above and compare the results for the cases of random and nonrandom designs.

5.7.1 Testing $\theta(t_0)=0$ versus $|\theta(t_0)|\geq 2c$

Given $t_0\in D$ and $c>0$, consider the hypotheses testing problem $\theta(t_0)=0$ versus $|\theta(t_0)|\geq 2c$. The risk $R_n^T(\rho_n)$ of test ρ_n is defined as above.

Proposition 5.9 *Let* $D=[0,1]$. *Let* c,M *and* t_0 *be such that* $[t_0-t^*,t_0+t^*]\subseteq[0,1]$, *where* $t^*=(c/M)^{1/\beta}$. *Let conditions* (P) *and* (Π) *hold and the function* $H_\gamma(s)$ *monotonously increase in* $s\geq 0$.
If $\Theta=\Sigma(\beta,M)$, *then*

$$\varliminf_{n\to\infty}\inf_{\rho_n}R_n^T(\rho_n)\geq T^*,$$

where

$$T^*=\inf_{\gamma\in[0,1]}\log\left(1+\int_{t_0-t^*}^{t_0+t^*}\left[H_\gamma(2(c-M|t-t_0|^\beta))-1\right]\pi(t)\,dt\right).$$

If $\Theta=\Sigma_0(\beta,M)$, *then*

$$\lim_{n\to\infty}\inf_{\rho_n}R_n^T(\rho_n)=T^*$$

and the tests $\rho_{n,\delta}^T$ *from (4.8) are nearly LD efficient, i.e.,*

$$\lim_{\delta\to 0}\varlimsup_{n\to\infty}R_n^T(\rho_{n,\delta}^T)=\lim_{\delta\to 0}\varliminf_{n\to\infty}R_n^T(\rho_{n,\delta}^T)=T^*.$$

Proof Theorem 4.3 reduces proof to calculating the value of T^* from (4.6). Using the result of Lemma 5.8 and proceeding in analogy with the case of deterministic design, we conclude that

$$\begin{aligned}&T^*=S(c-\theta^*,c+\theta^*)\\ =\;&\inf_{\gamma\in[0,1]}\log\left(\int_0^{t_0-t^*}\pi(t)\,dt+\int_{t_0-t^*}^{t_0+t^*}H_\gamma(2(c-M|t-t_0|^\beta))\pi(t)\,dt+\int_{t_0+t^*}^{1}\pi(t)\,dt\right).\end{aligned}$$

Now the claim follows by the equality $\int_D\pi(t)dt=1$. $\square$

5.7.2 Estimating $\theta(t_0)$

In estimating $\theta(t_0)$ the risk of estimator ρ_n is defined by

$$R_n^F(\rho_n) = \sup_{\theta \in \Sigma_0(\beta,M)} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta(t_0)| > c).$$

Proposition 5.10 *Let the conditions and notation of Proposition 5.9 hold. If $\Theta = \Sigma(\beta, M)$, then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) \geq F^*,$$

where

$$F^* = \inf_{\gamma\in[0,1]} \log \left(1 + \int_{t_0-t^*}^{t_0+t^*} \left[H_\gamma(2(c - M|t-t_0|^\beta)) - 1 \right] \pi(t)\, dt \right).$$

If $\Theta = \Sigma_0(\beta, M)$, then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) = F^*$$

and the interval-median estimators $\rho_{n,\delta}^F$ from (4.18) are nearly LD efficient, i.e.,

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R_n^F(\rho_{n,\delta}^F) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R_n^F(\rho_{n,\delta}^F) = F^*.$$

Proof Again it suffices to calculate the value of the asymptotic minimax risk F^* which is done as above. $\square$

Remark 5.13 *If we consider uniform design on $[0,1]$, i.e., take $\pi(t) = 1$, Jensen's inequality easily implies that the asymptotic minimax risk for regression with random design is not greater that the one for regression with deterministic design (see Subsection 5.5). This also follows from Lemma 5.5.*

Remark 5.14 *The problem of estimating $\theta(t_0)$ for uniform random design has been considered by Korostelev, 1995 who studies the double asymptotics as $n \to \infty$ and then $c \to 0$*

A Appendix

A.1 Proof of Lemma 2.5

Let $\{\mathbf{V}_\Lambda,\, \Lambda \in \mathcal{A}(\Theta)\}$ be a family of deviabilities satisfying property (S): for any $\Lambda \subset \Lambda'$ and $z_\Lambda \in S_\Lambda$

$$\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_{\Lambda'} \in \Pi^{-1}_{\Lambda'\Lambda} z_\Lambda} \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'})\,. \tag{A.1}$$

Recall that $\Pi^{-1}_{\Lambda'\Lambda} z_\Lambda = \{z_{\Lambda'} \in S_{\Lambda'} : \Pi_{\Lambda'\Lambda} z_{\Lambda'} = z_\Lambda\}$.

We define

$$\mathbf{V}_\Theta(z_\Theta) = \begin{cases} \inf_{\Lambda \in \mathcal{A}(\Theta)} \|\pi_\Lambda z_\Theta\|^{-1}_\Lambda \mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta)\,, & z_\Theta \in S_\Theta\,, \\ 0\,, & \text{otherwise,} \end{cases} \tag{A.2}$$

where we set $\mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta) = 1$ and $\|\pi_\Lambda z_\Theta\|^{-1}_\Lambda \mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta) = \infty$ if $\|\pi_\Lambda z_\Theta\|_\Lambda = 0$.

The functions $\|\pi_\Lambda z_\Theta\|^{-1}_\Lambda \mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta)\,, \Lambda \in \mathcal{A}(\Theta)$, are easily seen to be upper semicontinuous on S_Θ, so $(\mathbf{V}_\Theta(z_\Theta)\,, z_\Theta \in R^\Theta_+)$ is upper semicontinuous as the inf of a family of upper semicontinuous functions. Further, since, for every $z_\Theta \in S_\Theta$ and $\varepsilon > 0$, there exists $\Lambda \in \mathcal{A}(\Theta)$ such that $\|\pi_\Lambda z_\Theta\|_\Lambda > 1 - \varepsilon$ and since $\mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta) \leq 1$, we conclude that $\mathbf{V}_\Theta(z_\Theta) \leq 1$. Since (ii) obviously follows by (iii), we are left to prove (iii) and

$$\sup_{z_\Theta \in S_\Theta} \mathbf{V}_\Theta(z_\Theta) = 1. \tag{A.3}$$

We begin with (iii). Let us fix Λ and z_Λ. Definition (A.2) obviously implies that

$$\mathbf{V}_\Lambda(z_\Lambda) \geq \sup_{z_\Theta \in \Pi^{-1}_\Lambda z_\Lambda} \|\pi_\Lambda z_\Theta\|_\Lambda \mathbf{V}_\Theta(z_\Theta)\,.$$

So we need to prove that

$$\mathbf{V}_\Lambda(z_\Lambda) \leq \sup_{z_\Theta \in \Pi^{-1}_\Lambda z_\Lambda} \|\pi_\Lambda z_\Theta\|_\Lambda \mathbf{V}_\Theta(z_\Theta)\,. \tag{A.4}$$

We, first, note that (A.2) and (A.1) imply that

$$\mathbf{V}_\Theta(z_\Theta) = \inf_{\substack{\Lambda' \in \mathcal{A}(\Theta) \\ \Lambda' \supset \Lambda}} \|\pi'_\Lambda z_\Theta\|^{-1}_{\Lambda'} \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} z_\Theta)\,,\ z_\Theta \in S_\Theta\,. \tag{A.5}$$

Indeed, by (A.1), if $\Lambda \subset \Lambda' \in \mathcal{A}(\Theta)$ and $z_\Theta \in S_\Theta$ is such that $\|\pi_\Lambda z_\Theta\|_\Lambda > 0$, then

$$\mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta) \geq \|\pi_{\Lambda'\Lambda} \Pi_{\Lambda'} z_\Theta\|_\Lambda \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} z_\Theta)\,,$$

and hence, since $\pi_{\Lambda'\Lambda}\Pi_{\Lambda'} z_\Theta = \pi_\Lambda z_\Theta / \|\pi_{\Lambda'} z_\Theta\|_{\Lambda'}$,

$$\|\pi_{\Lambda'} z_\Theta\|^{-1}_{\Lambda'} \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} z_\Theta) \leq \|\pi_\Lambda z_\Theta\|^{-1}_\Lambda \mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta)\,,$$

which, in view of (A.2), proves (A.5).

Next, we obviously can assume that $a := \mathbf{V}_\Lambda(z_\Lambda) > 0$. For $\Lambda' \supset \Lambda\,, \Lambda' \in \mathcal{A}(\Theta)$, introduce the sets

$$A_{\Lambda'} = \{z_{\Lambda'} \in S_{\Lambda'} : \Pi_{\Lambda'\Lambda} z_{\Lambda'} = z_\Lambda \text{ and } \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}) = a\}. \tag{A.6}$$

We show that $A_{\Lambda'}$ is nonempty. Since $\mathbf{V}_{\Lambda'}(z_{\Lambda'}) \le 1$, the sup on the right of (A.1) can be taken over the set $\Pi^{-1}_{\Lambda'\Lambda} z_\Lambda \cap \{\|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \ge a/2\}$. This set is closed since the projection $\Pi_{\Lambda'\Lambda}$ is continuous on the set $\{z_{\Lambda'} : \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \ge a/2\}$. Since $\mathbf{V}_{\Lambda'}$ is a deviability, it attains sups on closed sets, so the sup on the right of (A.1) is attained which is equivalent to $A_{\Lambda'}$ being nonempty. Next, $A_{\Lambda'}$ is closed and hence compact since $\mathbf{V}_{\Lambda'}$ is upper semicontinuous and, by (A.1) and the definition of a, $\|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}) = a$ if and only if $\|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}) \ge a$.

Now we introduce for each $\Lambda' \in \mathcal{A}(\Theta)\,,\ \Lambda \subset \Lambda'$,

$$\mathbf{A}_{\Lambda'} = \{z_\Theta \in [0,1]^\Theta : \Pi_{\Lambda'} z_\Theta \in A_{\Lambda'} \text{ and } \|\pi_{\Lambda'} z_\Theta\|_{\Lambda'} \ge a\}.$$

These sets are easily seen to be nonempty (e.g., if $z_{\Lambda'} \in A_{\Lambda'}$, then $z_\Theta = (z_\theta, \theta \in \Theta)$ defined by $(z_\theta, \theta \in \Lambda') = z_{\Lambda'}$ and $z_\theta = 0, \theta \notin \Lambda'$, belongs to $\mathbf{A}_{\Lambda'}$) and compact for Tihonov topology on $[0,1]^\Theta$ (the latter is because $\Pi_{\Lambda'}$ is continuous on the set $\{z_\Theta : \|\pi_{\Lambda'} z_\Theta\|_{\Lambda'} \ge a\}$).

We next show that for any Λ' and Λ'' from $\mathcal{A}(\Theta)$ containing Λ, the sets $\mathbf{A}_{\Lambda'}$ and $\mathbf{A}_{\Lambda''}$ have nonempty intersection. Indeed, let $\Lambda''' = \Lambda' \cup \Lambda''$ and let $z_\Theta \in [0,1]^\Theta$ be such that $z_\Theta \in \mathbf{A}_{\Lambda'''}$ and $\|\pi_{\Lambda'''} z_\Theta\| = 1$ (such a z_Θ obviously exists). We prove that $z_\Theta \in \mathbf{A}_{\Lambda'}$ and $z_\Theta \in \mathbf{A}_{\Lambda''}$.

Denote $z_{\Lambda'''} = \Pi_{\Lambda'''} z_\Theta\,,\ z_{\Lambda'} = \Pi_{\Lambda'} z_\Theta$. First note that, since $z_{\Lambda'''} \in A_{\Lambda'''}$,

$$\Pi_{\Lambda'\Lambda} z_{\Lambda'} = \Pi_\Lambda z_\Theta = \Pi_{\Lambda'''\Lambda} z_{\Lambda'''} = z_\Lambda. \tag{A.7}$$

Then using also the equality $\Pi_{\Lambda'''\Lambda'} z_{\Lambda'''} = z_{\Lambda'}$, we have by (A.1) that

$$\mathbf{V}_\Lambda(z_\Lambda) \ \ge\ \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}), \tag{A.8}$$

$$\mathbf{V}_{\Lambda'}(z_{\Lambda'}) \ \ge\ \|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'} \mathbf{V}_{\Lambda'''}(z_{\Lambda'''}). \tag{A.9}$$

Next, by the definitions of $z_{\Lambda'''}$ and $z_{\Lambda'}$,

$$\|\pi_{\Lambda'''\Lambda} z_{\Lambda'''}\|_\Lambda = \|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'} \cdot \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda,$$

so that, by (A.8) and (A.9),

$$\mathbf{V}_\Lambda(z_\Lambda) \ge \|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'''} \cdot \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_{\Lambda'} \mathbf{V}_{\Lambda'''}(z_{\Lambda'''}) = \|\pi_{\Lambda'''\Lambda} z_{\Lambda'''}\| \mathbf{V}_{\Lambda'''}(z_{\Lambda'''}).$$

Since $z_{\Lambda'''} \in A_{\Lambda'''}$, we actually have equality here and hence in (A.8) and (A.9). The first of them and (A.7) prove that $z_{\Lambda'} \in A_{\Lambda'}$. Equalities in (A.8) and (A.9) together imply, since $\mathbf{V}_{\Lambda'''}(z_{\Lambda'''}) \le 1, \|\pi_{\Lambda'\Lambda} z_{\Lambda'\Lambda}\|_\Lambda \le 1$, that $\|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'} \ge \mathbf{V}_{\Lambda'}(z_{\Lambda'}) \ge \mathbf{V}_\Lambda(z_\Lambda) = a$; since also $\|\pi_{\Lambda'''} z_\Theta\|_{\Lambda'''} = 1$, we get

$$\|\pi_{\Lambda'} z_\Theta\|_{\Lambda'} = \|\pi_{\Lambda'''} z_\Theta\|_{\Lambda'''} \cdot \|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'} \ge a.$$

This concludes the proof of the inclusion $z_\Theta \in \mathbf{A}_{\Lambda'}$. The inclusion $z_\Theta \in \mathbf{A}_{\Lambda''}$ is proved similarly.

Thus the finite intersections of the compacts $\mathbf{A}_{\Lambda'}, \Lambda' \supset \Lambda$, are nonempty, hence $\cap_{\Lambda' \supset \Lambda} \mathbf{A}_{\Lambda'} \neq \emptyset$. It remains to check that, for any z_Θ from this intersection,

$$\Pi_\Lambda z_\Theta = z_\Lambda \tag{A.10}$$

and

$$\mathbf{V}_\Theta(z_\Theta) \geq \|\pi_\Lambda z_\Theta\|_\Lambda^{-1} \mathbf{V}_\Lambda(z_\Lambda) \tag{A.11}$$

which obviously yields (A.4). Let $\Lambda' \in \mathcal{A}(\Theta)$ with $\Lambda \subset \Lambda'$. Since $\Pi_{\Lambda'} z_\Theta \in A_{\Lambda'}$, it follows that $\Pi_\Lambda z_\Theta = \Pi_{\Lambda'\Lambda} \Pi_{\Lambda'} z_\Theta = z_\Lambda$, checking (A.10), also

$$\mathbf{V}_\Lambda(z_\Lambda) = a = \|\pi_{\Lambda'\Lambda} \Pi_{\Lambda'} z_\Theta\|_\Lambda \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} z_\Theta) = \frac{\|\pi_\Lambda z_\Theta\|_\Lambda}{\|\pi_{\Lambda'} z_\Theta\|_{\Lambda'}} \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} z_\Theta),$$

so

$$\|\pi_\Lambda z_\Theta\|_\Lambda^{-1} \mathbf{V}_\Lambda(z_\Lambda) = \|\pi_{\Lambda'} z_\Theta\|_{\Lambda'}^{-1} \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} z_\Theta).$$

In view of (A.5), this implies (A.11) and (A.4) follows.

Finally, according to *(iii)*,

$$1 = \sup_{z_\Lambda \in S_\Lambda} \mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta \in S_\Theta} \|\pi_\Lambda z_\Theta\|_\Lambda \mathbf{V}_\Theta(z_\Theta) \leq \sup_{z_\Theta \in S_\Theta} \mathbf{V}_\Theta(z_\Theta),$$

proving (A.3). □

Remark A.1 *It is not difficult to see that (A.2) is equivalent to*

$$\mathbf{V}_\Theta(z_\Theta) = \lim_{\Lambda \in \mathcal{A}(\Theta)} \mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta), \; z_\Theta \in S_\Theta,$$

where the limit is with respect to the partial ordering by inclusion: $\Lambda \leq \Lambda'$ if $\Lambda \subset \Lambda'$.

A.2 A minimax theorem for non-level compact loss functions

This subsection contains a minimax theorem for generalised risks and non-level compact loss functions. We assume the setting described at the beginning of Section 3 and start by introducing an extension of the space of decisions, cf. Strasser, 1985.

Denote by $\mathcal{C}_+(\mathcal{D})$ the set of all nonnegative, bounded continuous functions on $\mathcal{D}$, and let $\mathbf{B}(\mathcal{D})$ be the set of all functionals $b : \mathcal{C}_+(\mathcal{D}) \to R_+$ with the following properties:

(1) $b(\mathbf{0}) = 0, b(\mathbf{1}) = 1$, where $\mathbf{0}$ (respectively, $\mathbf{1}$) denotes the element of $\mathcal{C}_+(\mathcal{D})$ identically equal to 0 (respectively, 1);

(2) $b(f) \leq b(g)$ if $f \leq g$, $f, g \in \mathcal{C}_+(\mathcal{D})$;

(3) $b(\lambda f) = \lambda b(f)$, $f \in \mathcal{C}_+(\mathcal{D}), \lambda \in R_+$;

(4) $b(f + g) \leq b(f) + b(g)$, $f, g \in \mathcal{C}_+(\mathcal{D})$.

Also let $\mathbf{B}_1(\mathcal{D})$ be the subset of those $b \in \mathbf{B}(\mathcal{D})$ for which, in addition,

(5) $b(f \vee g) = b(f) \vee b(g),\ f, g \in \mathcal{C}_+(\mathcal{D}),$

where $f \vee g$ denotes the maximum of f and g.

We endow $\mathbf{B}(\mathcal{D})$ with weak topology which is the topology induced by Tihonov (product) topology on $R_+^{\mathcal{C}_+(\mathcal{D})}$, i.e., a net $\{b_\sigma, \sigma \in \Sigma\}$, where Σ is a directed set, of elements of $\mathbf{B}(\mathcal{D})$ converges to $b \in \mathbf{B}(\mathcal{D})$ if, for all $f \in \mathcal{C}_+(\mathcal{D})$, $\lim_{\sigma\in\Sigma} b_\sigma(f) = b(f)$.

With each $r \in \mathcal{D}$, we associate the element b_r of $\mathbf{B}_1(\mathcal{D})$ defined by

$$b_r(f) = f(r),\ f \in \mathcal{C}_+(\mathcal{D}). \tag{A.12}$$

We extend the domain of the functionals b to the set $\underline{\mathcal{C}}_+(\mathcal{D})$ of lower semicontinuous nonnegative functions on $\mathcal{D}$ by letting

$$b(g) = \sup\{b(f):\ f \le g, f \in \mathcal{C}_+(\mathcal{D})\},\ g \in \underline{\mathcal{C}}_+(\mathcal{D}). \tag{A.13}$$

It is easily seen that the map $b \to b(g)$, for any $g \in \underline{\mathcal{C}}_+(\mathcal{D})$, is lower semicontinuous on $\mathbf{B}(\mathcal{D})$. Note also that, under extension (A.13), equality (A.12) carries over to functions g from $\underline{\mathcal{C}}_+(\mathcal{D})$ if and only if $g = \sup\{f : f \le g, f \in \mathcal{C}_+(\mathcal{D})\}$, which holds in particular if $\mathcal{D}$ is locally compact. Generally, however, $b_r(g) \le g(r)$.

Finally, denote by $\mathcal{B}_n$ the set of all random elements on $(\Omega_n, \mathcal{F}_n)$ with values in $\mathbf{B}(\mathcal{D})$. We call the elements of $\mathcal{B}_n$ generalised decision functions (or generalised decisions). Note that if $\rho_n \in \mathcal{R}_n$, then $b_{\rho_n} \in \mathcal{B}_n$.

Given loss functions $W_\theta, \theta \in \Theta$, the LD risk $B_n(\beta_n)$ of a generalised decision $\beta_n \in \mathcal{B}_n$ in the experiment $\mathcal{E}_n = (\Omega_n, \mathcal{F}_n; P_{n,\theta}, \theta \in \Theta)$ is defined by

$$B_n(\beta_n) = \sup_{\theta\in\Theta} E_{n,\theta}^{1/n} \beta_n(W_\theta^n). \tag{A.14}$$

Theorem A.1 *Let $\{\mathcal{E}_n, n \ge 1\}$ satisfy the LDP. Then*

$$\varliminf_{n\to\infty} \inf_{\beta_n\in\mathcal{B}_n} B_n(\beta_n) \ge B^*,$$

where

$$B^* = \sup_{z_\Theta\in R_+^\Theta} \inf_{b\in\mathbf{B}_1(\mathcal{D})} \sup_{\theta\in\Theta} b(W_\theta) z_\theta \mathbf{V}_\Theta(z_\Theta).$$

For a proof, we need to study properties of $\mathbf{B}(\mathcal{D})$ and $\mathbf{B}_1(\mathcal{D})$.

Lemma A.1 *For any finite number of functions $f_1, f_2, \ldots, f_k \in \mathcal{C}_+(\mathcal{D})$ and any sequence $\{b_n, n \ge 1\}$ of elements of $\mathbf{B}(\mathcal{D})$, there exists $b \in \mathbf{B}_1(\mathcal{D})$ such that $b(f_i)$ is an accumulation point of the sequence $\{b_n^{1/n}(f_i^n), n \ge 1\}$ for $i = 1, \ldots, k$.*

Proof Let $||\cdot||$ denote the uniform norm on $\mathcal{C}_+(\mathcal{D})$. Define $\mathcal{C}_{1,+}(\mathcal{D})$ as the subset of $\mathcal{C}_+(\mathcal{D})$ of functions f with $||f|| \le 1$. Introduce the functionals $\overline{b}_n(f) = b_n^{1/n}(f^n), f \in \mathcal{C}_{1,+}(\mathcal{D})$. Then the set $B = \{\overline{b}_n, n \ge 1\}$ belongs to the set $[0,1]^{\mathcal{C}_{1,+}(\mathcal{D})}$. By Tihonov's theorem, $[0,1]^{\mathcal{C}_{1,+}(\mathcal{D})}$ submitted with product topology is compact, and hence B is relatively compact. Let $\tilde{b}$ denote some its accumulation point. Since $b_n \in \mathbf{B}(\mathcal{D})$ and by the definition of product topology, $\tilde{b}$ has properties (1), (2) and (4) of $\mathbf{B}(\mathcal{D})$. We extend $\tilde{b}$ to $\mathcal{C}_+(\mathcal{D})$ by letting $\tilde{b}(\lambda f) = \lambda \tilde{b}(f), \lambda > 0, f \in \mathcal{C}_{1,+}(\mathcal{D})$. Then $\tilde{b} \in \mathbf{B}(\mathcal{D})$. Also, since the topology on $\mathbf{B}(\mathcal{D})$ is the restriction of product topology on $R_+^{\mathcal{C}_+(\mathcal{D})}$, $\tilde{b}$ is an accumulation point of $\{\overline{b}_n, n \ge 1\}$, where $\overline{b}_n$ are extended to $\mathcal{C}_+(\mathcal{D})$ by $\overline{b}_n(\lambda f) = \lambda \overline{b}_n(f), \lambda > 0, f \in \mathcal{C}_{1,+}(\mathcal{D})$. This implies, by the definition of the $\overline{b}_n$, that $\tilde{b}(f_i)$ is an accumulation point of $\{b_n^{1/n}(f_i^n), n \ge 1\}$ for $i = 1, \dots, k$.

We end the proof by showing that $\tilde{b} \in \mathbf{B}_1(\mathcal{D})$. Let $f, g \in \mathcal{C}_+(\mathcal{D})$. Then, since $\tilde{b}$ is an accumulation point of $\{\overline{b}_n, n \ge 1\}$, $\tilde{b}(f)$, $\tilde{b}(g)$ and $\tilde{b}(f \vee g)$ are the respective accumulation points of $\{\overline{b}_n(f), n \ge 1\}$, $\{\overline{b}_n(g), n \ge 1\}$ and $\{\overline{b}_n(f \vee g), n \ge 1\}$. Hence, by the definition of the $\overline{b}_n$, for a subsequence (n'), $b_{n'}^{1/n'}(f^{n'}) \to \tilde{b}(f)$, $b_{n'}^{1/n'}(g^{n'}) \to \tilde{b}(g)$ and $b_{n'}^{1/n'}((f \vee g)^{n'}) \to \tilde{b}(f \vee g)$. By properties (2) and (4) of $\mathbf{B}(\mathcal{D})$,

$$b_n^{1/n}(f^n) \vee b_n^{1/n}(g^n) \le b_n^{1/n}((f \vee g)^n) \le 2^{1/n} \left[b_n^{1/n}(f^n) \vee b_n^{1/n}(g^n) \right],$$

and we conclude that $\tilde{b}(f \vee g) = \tilde{b}(f) \vee \tilde{b}(g)$. □

Lemma A.2 *The set* $\mathbf{B}_1(\mathcal{D})$ *is compact.*

Proof An argument similar to the one used in Lemma A.1 shows that the set of functionals $\{(b(f), f \in \mathcal{C}_{1,+}(\mathcal{D})), b \in \mathbf{B}_1(\mathcal{D})\}$ is closed in $[0,1]^{\mathcal{C}_{1,+}(\mathcal{D})}$ and hence it is compact which obviously is equivalent to the compactness of $\mathbf{B}_1(\mathcal{D})$. □

The next lemma is motivated by and extends Aubin, 1984, Proposition 8.2.

Lemma A.3 *Let T be an arbitrary set and let U be a topological space. Assume that a real-valued function $g(t,u), t \in T, u \in U$, has the properties:*

(a) for any $t \in T$, $g(t,u)$ is level compact in $u \in U$,

(b) for any $t_1, t_2 \in T$, there exists $t_3 \in T$ such that $g(t_3,u) \ge g(t_1,u) \vee g(t_2,u)$ for all $u \in U$.

Then

$$\sup_{t \in T} \inf_{u \in U} g(t,u) = \inf_{u \in U} \sup_{t \in T} g(t,u).$$

Remark A.2 *Condition (a) holds if $g(t,u)$ is lower semicontinuous in u and U is a compact topological space.*

Remark A.3 *If T is a directed set, condition (b) holds if $g(t,u)$ is increasing in t for all u, i.e., $g(t_1,u) \le g(t_2,u)$, $u \in U$, each time as $t_1 \le t_2$ (the latter relation is with respect to the order on T).*

Proof We proceed as in Aubin, 1984. Pick $\alpha > \sup_{t\in T}\inf_{u\in U} g(t,u)$. We need to prove that

$$\alpha \geq \inf_{u\in U}\sup_{t\in T} g(t,u). \tag{A.15}$$

Let $T_0 = \{t \in T : \sup_{u\in U} g(t,u) > \alpha\}$. If T_0 is empty, the proof is over. So we assume that $T_0 \neq \emptyset$. By the definition of α, the sets $A_t = \{u \in U : g(t,u) \leq \alpha\}$ are nonempty for all $t \in T$, and they are, moreover, compact for all $t \in T_0$, since $g(t,u), u \in U$, are level compact. Condition (b) implies that, for every $t_1, t_2 \in T$, there exists $t_3 \in T$ such that $A_{t_1} \bigcap A_{t_2} \supset A_{t_3} \neq \emptyset$, which shows that the finite intersections of the compacts $A_t, t \in T_0$, are nonempty, and hence $\bigcap_{t\in T_0} A_t \neq \emptyset$. The latter is equivalent to

$$\alpha \geq \inf_{u\in U}\sup_{t\in T_0} g(t,u).$$

Since by the definition of T_0, $\alpha \geq \sup_{t\in T\setminus T_0} g(t,u)$, $u \in U$, (A.15) is proved. □

Proof of Theorem A.1 We need to prove that, for an arbitrary sequence $\beta_n, n \geq 1$, of generalised decisions,

$$\underline{\lim}_{n\to\infty}\; B_n(\beta_n) \geq B^*. \tag{A.16}$$

The argument is similar to the one in the proof of Theorem 3.1. Let $f_\theta(r)\,, \theta \in \Theta$, be some nonnegative, bounded and continuous in $r \in \mathcal{D}$ functions. Fix a nonempty $\Lambda \in \mathcal{A}(\Theta)$. We have, by the definition of $\mathbf{Z}_{n,\Lambda}$ (see (2.13)), that

$$\begin{aligned} \underline{\lim}_{n\to\infty}\; \sup_{\theta\in\Lambda} E_{n,\theta}^{1/n}\beta_n(f_\theta^n) &= \underline{\lim}_{n\to\infty}\; \sup_{\theta\in\Lambda} E_{n,\Lambda}^{1/n}\beta_n(f_\theta^n)\mathbf{Z}_{n,\theta;\Lambda}^n \\ \geq\; \underline{\lim}_{n\to\infty}\; \left[\frac{1}{|\Lambda|}E_{n,\Lambda}\sum_{\theta\in\Lambda}\beta_n(f_\theta^n)\mathbf{Z}_{n,\theta;\Lambda}^n\right]^{1/n} &\geq\; \underline{\lim}_{n\to\infty}\; E_{n,\Lambda}^{1/n}\sup_{\theta\in\Lambda}\beta_n(f_\theta^n)\mathbf{Z}_{n,\theta;\Lambda}^n \\ &\geq\; \underline{\lim}_{n\to\infty}\; E_{n,\Lambda}^{1/n}u_n^n(\mathbf{Z}_{n,\Lambda}), \end{aligned} \tag{A.17}$$

where

$$u_n(z_\Lambda) = \inf_{b\in\mathbf{B}(\mathcal{D})}\sup_{\theta\in\Lambda} b^{1/n}(f_\theta^n)z_\theta, \quad z_\Lambda = (z_\theta, \theta \in \Lambda) \in R_+^\Lambda. \tag{A.18}$$

Note that the $u_n(z_\Lambda)\,, n = 1, 2, \ldots,$ are upper semicontinuous (recall that Λ is finite) and hence measurable, so that the expectations on the right most side of (A.17) are well defined.

Introduce

$$u(z_\Lambda) = \inf_{b\in\mathbf{B}_1(\mathcal{D})}\sup_{\theta\in\Lambda} b(f_\theta)z_\theta, \quad z_\Lambda \in R_+^\Lambda. \tag{A.19}$$

We prove that

$$\underline{\lim}_{n\to\infty}\; u_n(z_\Lambda(n)) \geq u(z_\Lambda), \quad z_\Lambda \in R_+^\Lambda, \tag{A.20}$$

for any sequence $z_\Lambda(n) \to z_\Lambda$.

Let $b_n \in \mathbf{B}(\mathcal{D})$ be such that

$$\underline{\lim}_{n\to\infty}\; u_n(z_\Lambda(n)) = \underline{\lim}_{n\to\infty}\; \sup_{\theta\in\Lambda} b_n^{1/n}(f_\theta^n)z_\theta(n).$$

By Lemma A.1 and since Λ is finite, there exists $\tilde b \in \mathbf{B}_1(\mathcal{D})$ such that $\tilde b(f_\theta)$ is an accumulation point of $\{b_n^{1/n}(f_\theta^n), n \geq 1\}$ for all $\theta \in \Lambda$. Therefore we have that, for a subsequence (n'),

$$\lim_{n'} b_{n'}^{1/n'}(f_\theta^{n'}) = \tilde b(f_\theta), \quad \theta \in \Lambda,$$
$$\lim_{n'} \sup_{\theta\in\Lambda} b_{n'}^{1/n'}(f_\theta^{n'}) z_\theta(n') = \varliminf_{n\to\infty} \sup_{\theta\in\Lambda} b_n^{1/n}(f_\theta^n) z_\theta(n).$$

Since Λ is finite and $z_\Lambda(n') \to z_\Lambda$, we conclude that

$$\varliminf_{n\to\infty} \sup_{\theta\in\Lambda} b_n^{1/n}(f_\theta^n) z_\theta(n) = \sup_{\theta\in\Lambda} \tilde b(f_\theta) z_\theta$$

which, in view of (A.19), proves (A.20).

By (A.20) and the LD convergence of $\{\mathcal{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}), n \geq 1\}$ to $\mathbf{V}_\Lambda$, we have that (see Varadhan, 1966; Chaganty, 1993; Puhalskii, 1995a)

$$\varliminf_{n\to\infty} E_{n,\Lambda}^{1/n} u_n^n(\mathbf{Z}_{n,\Lambda}) \geq \sup_{z_\Lambda \in R_+^\Lambda} u(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda). \tag{A.21}$$

Since by (A.19) $u \in \mathcal{H}_\Lambda$, property (ii) of $\mathbf{V}_\Theta$ in Lemma 2.5 yields

$$\sup_{z_\Lambda \in R_+^\Lambda} u(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta \in R_+^\Theta} u(\pi_\Lambda z_\Theta)\mathbf{V}_\Theta(z_\Theta).$$

Relations (A.17) and (A.21) imply then that

$$\varliminf_{n\to\infty} \sup_{\theta\in\Lambda} E_{n,\theta}^{1/n} \beta_n(f_\theta^n) \geq \sup_{z_\Theta \in R_+^\Theta} u(\pi_\Lambda z_\theta)\mathbf{V}_\Theta(z_\Theta),$$

so, by the definition of the function u in (A.19),

$$\varliminf_{n\to\infty} \sup_{\theta\in\Lambda} E_{n,\theta}^{1/n} \beta_n(f_\theta^n) \geq \sup_{z_\Theta \in R_+^\Theta} \inf_{b\in\mathbf{B}_1(\mathcal{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta \mathbf{V}_\Theta(z_\Theta)$$

and hence, since $\Lambda \in \mathcal{A}(\Theta)$ and $f_\theta \leq W_\Theta, f_\theta \in \mathcal{C}_+(\mathcal{D})$, $\theta \in \Theta$, are otherwise arbitrary,

$$\varliminf_{n\to\infty} \sup_{\theta\in\Theta} E_{n,\theta}^{1/n} \beta_n(W_\theta^n) \geq \sup_{z_\Theta \in R_+^\Theta} \sup_{\substack{\Lambda\in\mathcal{A}(\Theta)\\ f_\Theta \in C_W}} \inf_{b\in\mathbf{B}_1(\mathcal{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta \mathbf{V}_\Theta(z_\Theta),$$

where $C_W = \{f_\Theta = (f_\theta, \theta\in\Theta) \in \mathcal{C}_+(\mathcal{D})^\Theta : \ f_\theta \leq W_\theta, \theta \in \Theta\}$. Thus (A.16) and the theorem would follow if, for every $z_\Theta = (z_\theta, \theta\in\Theta) \in R_+^\Theta$,

$$\sup_{\substack{\Lambda\in\mathcal{A}(\Theta)\\ f_\Theta \in C_W}} \inf_{b\in\mathbf{B}_1(\mathcal{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta = \inf_{b\in\mathbf{B}_1(\mathcal{D})} \sup_{\theta\in\Theta} b(W_\theta) z_\theta. \tag{A.22}$$

Fixing z_Θ, introduce, for $\Lambda \in \mathcal{A}(\Theta), f_\Theta \in \mathcal{C}_+(\mathcal{D})^\Theta, b \in \mathbf{B}_1(\mathcal{D})$,

$$g((\Lambda, f_\Theta), b) = \sup_{\theta\in\Lambda} b(f_\theta) z_\theta.$$

We check that $g((\Lambda, f_\Theta), b)$ satisfies the conditions of Lemma A.3. Submit the set $\mathcal{A}(\Theta) \times C_W$ with natural order: $(\Lambda, f_\Theta) \le (\Lambda', f'_\Theta)$ if $\Lambda \subset \Lambda'$ and $f_\theta \le f'_\theta, \theta \in \Theta$. It is easily seen that $\mathcal{A}(\Theta) \times C_W$ is a directed set and $g((\Lambda, f_\Theta), b)$ is increasing for each b. Also $g((\Lambda, f_\Theta), b)$ is continuous in b for each (Λ, f_Θ) by the definition of topology on $\mathbf{B}(\mathcal{D})$ and since Λ is finite; since $\mathbf{B}_1(\mathcal{D})$ is compact by Lemma A.2, $g((\Lambda, f_\Theta), b)$ is level compact in b. Thus by Lemma A.3,

$$\sup_{(\Lambda, f_\Theta) \in \mathcal{A}(\Theta) \times C_W} \inf_{b \in \mathbf{B}_1(\mathcal{D})} g((\Lambda, f_\Theta), b) = \inf_{b \in \mathbf{B}_1(\mathcal{D})} \sup_{(\Lambda, f_\Theta) \in \mathcal{A}(\Theta) \times C_W} g((\Lambda, f_\Theta), b).$$

Recalling the definition of g and using that by (A.13)

$$b(W_\theta) = \sup\{b(f_\theta) : \ f_\theta \le W_\theta, f_\theta \in \mathcal{C}_+(\mathcal{D})\}, \quad \theta \in \Theta,$$

we get (A.22). $\square$

It is interesting to find out how Theorem A.1 relates to Theorem 3.1. Above definitions easily imply that $\inf_{\rho_n \in \mathcal{R}_n} R_n(\rho_n) \ge \inf_{\beta_n \in \mathcal{B}_n} B_n(\beta_n)$ and $R^* \ge B^*$. The next lemma shows, in particular, that if $\mathcal{D}$ is locally compact, then Theorem 3.1 is a consequence of Theorem A.1.

Lemma A.4 *If the loss functions W_θ are such that*

$$W_\theta = \sup\{f_\theta : \ f_\theta \le W_\theta, f_\theta \in \mathcal{C}_+(\mathcal{D}), f_\theta \text{ are level compact } \}, \theta \in \Theta,$$

then

$$R^* = B^*.$$

Remark A.4 *The conditions of the lemma hold if the W_θ are level compact and $\mathcal{D}$ is locally compact (cf. Strasser, 1985, Theorem 6.4).*

A proof is preceded by two lemmas. We first derive a maxitive analogue of the partition of the unity (cf. again Strasser, 1985, Lemma 6.6).

Lemma A.5 *Let $f_1, \ldots, f_k \in \mathcal{C}_+(\mathcal{D})$. For any $\varepsilon > 0$, there exist $h_1, \ldots, h_m \in \mathcal{C}_+(\mathcal{D})$ with the properties:*

1^0 $\max_{1 \le j \le m} h_j(r) = 1, r \in \mathcal{D}$,

2^0 *for every $j = 1, \ldots, m$, $\max_{1 \le i \le k} |f_i(r_1) - f_i(r_2)| \le \varepsilon$ for any r_1 and r_2 such that $h_j(r_1) > 0, h_j(r_2) > 0$.*

Proof The argument is similar to that in Strasser, 1985. Assume first that $k = 1$ and $\sup_{r \in \mathcal{D}} f_1(r) = 1$. Choose m such that $3/m \le \varepsilon$ and define, for $x \ge 0$,

$$g_j(x) = (x - (j-2))^+ \wedge (j + 1 - x)^+ \wedge 1, \ 1 \le j \le m.$$

Let

$$h_j(r) = g_j(m f_1(r)), \ 1 \le j \le m, r \in \mathcal{D}.$$

It is readily seen, since $g_j(x) = 1$ if $j-1 \le x \le j$ and $\sup_{r\in\mathcal{D}} mf_1(r) = m$, that $\max_{1\le j\le m} h_j(r) = 1, r \in \mathcal{D}$.

Next, since, for $j = 1,\dots,m$, $g_j(x) = 0$ if $x \notin [(j-2)^+, j+1]$, we have that if $h_j(r_1) > 0$ and $h_j(r_2) > 0$, then $|mf_1(r_1) - mf_1(r_2)| \le 3$, i.e., $|f_1(r_1) - f_1(r_2)| \le 3/m \le \varepsilon$ as required.

Now if $\sup_{r\in\mathcal{D}} f_1(r) = a > 0$, then the h_j chosen as above for f_1/a and ε/a satisfy 1^0 and 2^0.

Finally, if $k > 1$, choose, for each $i = 1,\dots,k$, functions $h_{i,j}, 1 \le j \le m_i$, which satisfy 1^0 and 2^0. Then the functions

$$h_{j_1,\dots,j_k}(r) = \prod_{i=1}^{k} h_{i,j_i}(r),\ 1 \le j_i \le m_i, r \in \mathcal{D},$$

meet the required for all i with $m = m_1 \dots m_k$. $\square$

Denote by T_1 the set of nonnegative (upper semicontinuous) functions of finite support $(t(r), r \in \mathcal{D})$ such that $\sup_{r\in\mathcal{D}} t(r) = 1$. Define $\mathbf{B}_2(\mathcal{D})$ as the set of those $b \in \mathbf{B}_1(\mathcal{D})$ which can be represented as $b(f) = \sup_{r\in\mathcal{D}} f(r)t(r)$, $f \in \mathcal{C}_+(\mathcal{D})$, for some $(t(r), r \in \mathcal{D}) \in T_1$. The next lemma parallels Strasser, 1985, Theorem 42.5.

Lemma A.6 *The set $\mathbf{B}_2(\mathcal{D})$ is dense in $\mathbf{B}_1(\mathcal{D})$ for the weak topology.*

Proof We proceed as in the proof of Strasser, 1985, Theorem 42.5. Fix $b \in \mathbf{B}_1(\mathcal{D})$ and $f_1,\dots,f_k \in \mathcal{C}_+(\mathcal{D})$. We have to check that for any $\varepsilon > 0$ there exists $\tilde{b} \in \mathbf{B}_2(\mathcal{D})$ such that $|b(f_i) - \tilde{b}(f_i)| \le \varepsilon$, $1 \le i \le k$.

Let functions h_j, $1 \le j \le m$, be as in Lemma A.5. Obviously we can assume that they are not identically equal to 0. For each $j = 1,\dots,m$, choose r_j such that $h_j(r_j) > 0$. By the definition of the h_j,

$$|f_i(r)h_j(r) - f_i(r_j)h_j(r)| \le \varepsilon,\ 1 \le i \le k, r \in \mathcal{D},$$

on the one hand, and

$$f_i(r) = \max_{1\le j\le m} f_i(r)h_j(r),\ 1 \le i \le k, r \in \mathcal{D},$$

on the other hand. Hence

$$|f_i(r) - \max_{1\le j\le m} f_i(r_j)h_j(r)| \le \max_{1\le j\le m} |f_i(r)h_j(r) - f_i(r_j)h_j(r)| \le \varepsilon,$$
$$1 \le i \le k, r \in \mathcal{D}.$$

Properties (1), (3) and (4) of $\mathbf{B}(\mathcal{D})$ then yield

$$|b(f_i) - b(\max_{1\le j\le m} f_i(r_j)h_j)| \le \varepsilon,\ 1 \le i \le k. \tag{A.23}$$

Now since $b \in \mathbf{B}_1(\mathcal{D})$ and by property (3) again,

$$b(\max_{1\le j\le m} f_i(r_j)h_j) = \max_{1\le j\le m} f_i(r_j)b(h_j),\ 1 \le i \le k. \tag{A.24}$$

Define

$$t(r) = \begin{cases} \max_{l: r_l = r_j} b(h_l), & \text{if } r = r_j \text{ for some } j = 1, \ldots, m, \\ 0, & \text{otherwise,} \end{cases}$$

and let

$$\tilde{b}(f) = \sup_{r \in \mathcal{D}} f(r) t(r), \ f \in \mathcal{C}_+(\mathcal{D}).$$

By properties (1) and (5) of $\mathbf{B}_1(\mathcal{D})$, and the choice of the h_j,

$$\sup_{r \in \mathcal{D}} t(r) = \max_{1 \le j \le m} b(h_j) = b(\max_{1 \le j \le m} h_j) = b(\mathbf{1}) = 1,$$

so $(t(r)) \in T_1$.

Also by the definitions of $t(r)$ and $\tilde{b}$, the right hand side of (A.24) equals $\tilde{b}(f_i)$, and (A.24) and (A.23) yield the required. $\square$

Proof of Lemma A.4 Since $R^* \ge B^*$, we prove the opposite inequality. Let $f_\theta, \theta \in \Theta$, be level compact, belong to $\mathcal{C}_+(\mathcal{D})$ and $f_\theta \le W_\theta, \theta \in \Theta$. By the definition of B^*,

$$B^* \ge \sup_{z_\Theta \in R_+^\Theta} \inf_{b \in \mathbf{B}_1(\mathcal{D})} \sup_{\theta \in \Lambda} b(f_\theta) z_\theta \mathbf{V}_\Theta(z_\Theta), \ \Lambda \in \mathcal{A}(\Theta). \tag{A.25}$$

By Lemma A.6, for $z_\Theta \in R_+^\Theta, \Lambda \in \mathcal{A}(\Theta)$,

$$\begin{aligned} &\inf_{b \in \mathbf{B}_1(\mathcal{D})} \sup_{\theta \in \Lambda} b(f_\theta) z_\theta = \inf_{b \in \mathbf{B}_2(\mathcal{D})} \sup_{\theta \in \Lambda} b(f_\theta) z_\theta \\ &= \inf_{(t(r)) \in T_1} \sup_{r \in \mathcal{D}} \sup_{\theta \in \Lambda} t(r) f_\theta(r) z_\theta = \inf_{r \in \mathcal{D}} \sup_{\theta \in \Lambda} f_\theta(r) z_\theta. \end{aligned}$$

Since the f_θ are level compact, an application of Lemma A.3 shows, in analogy with the end of the proof of Theorem A.1, that the sup of the latter quantity over the f_θ and $\Lambda \in \mathcal{A}(\Theta)$ equals $\inf_{r \in \mathcal{D}} \sup_{\theta \in \Theta} W_\theta(r) z_\theta$ which by (A.25) proves that $B^* \ge R^*$. $\square$

References

AUBIN, J.-P. (1984). *L'Analyse non Linéaire et ses Motivations Économiques.* Masson.

AUBIN, J.-P. AND EKELAND, I. (1984). *Applied Nonlinear Analysis.* Wiley.

BAHADUR, R. (1960). On the asymptotic efficiency of tests and estimators. *Sankhya*, 22:229–252.

BAIOCCHI, C. AND CAPELO, A. (1984). *Variational and Quasivariational Inequalities.* Wiley.

BASU, D. (1956). On the concept of asymptotic efficiency. *Sankhya*, 17:193 –196.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley.

BOROVKOV, A. AND MOGULSKII, A. (1992a). Large deviations and statistical invariance principle. *Theory Probab. Appl.*, 37:11–18.

BOROVKOV, A. AND MOGULSKII, A. (1992b). Large deviations and testing of statistical hypotheses. *Proc. Inst. Math. Russian Acad. Sci., Siberian Division*, 19. English transl.: Siber. Adv. Math. , V. 2, 1992 N.3,4. V. 3 ,1993 N.1,2.

BRYC, W. (1990). Large deviations by the asymptotic value method. In Pinsky, M. A., editor, *Diffusion processes and related problems in analysis*, volume 1, pages 447–472. Birkhauser.

CHAGANTY, N. (1993). Large deviations for joint distributions and statistical applications. Technical Report TR93-2, Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23:497–507.

DAWSON, D. AND GÄRTNER, J. (1987). Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochast. Stochast. Rep.*, 20:247–308.

DEUSCHEL, J. AND STROOCK, D. (1989). *Large Deviations.* Academic Press.

ENGELKING, R. (1977). *General Topology.* Państwowe Wydawnictwo Naukowe.

FREIDLIN, M. AND WENTZELL, D. (1984). *Random Perturbations of Dynamical Systems.* Nauka. in Russian; English translation: Springer, 1984.

FU, J. (1982). Large sample point estimation: a large deviation theory approach. *Ann. Statist.*, 10:762–771.

IBRAGIMOV, I. AND KHASMINSKII, R. (1981). *Statistical Estimation: Asymptotic Theory.* Springer. New York, etc.

IBRAGIMOV, I. AND RADAVICHIUS, M. (1981). Probability of large deviations for the maximum likelihood estimator. *Sov. Math. Dokl.*, 23(2):403–406.

KALLENBERG, W. (1981). Bahadur deficiency of likelihood ratio tests in exponential families. *J. Multivariate Anal.*, 11:506 – 53.

KALLENBERG, W. (1983). Intermediate efficiency, theory and examples. *Ann. Statist.*, 11:170–182.

KOROSTELEV, A. (1993). Minimaxity criterion in nonparametric regression based on large deviations probabilities. *Ann. Statist.* to appear.

KOROSTELEV, A. (1995). Minimax large deviations risk in change-point problems. *Stochastic Process and Their Appl.* submitted.

KOROSTELEV, A. AND LEONOV, S. (1995). Minimax Bahadur efficiency for small confidence intervals. Discussion paper 37, SFB 373, Humboldt University, Berlin.

KOROSTELEV, A. AND SPOKOINY, V. (1995). Exact asymptotics of minimax Bahadur risk in Lipschitz regression. *Statistics.* to appear.

KRASNOSELSKII, M. AND RUTICKII, Y. (1961). *Convex functions and Orlicz spaces.* Noordhoff.

KULLBACK, S. (1959). *Information Theory and Statistics.* Wiley, New York.

LECAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. Calif. Publ. Statist.*, 3:27–98.

LECAM, L. (1972). Limits of experiments. *Proc. 6th Berkeley Symp. Math. Stat. Prob.*, 1:245–261.

LECAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York etc.

LYNCH, J. AND SETHURAMAN, J. (1987). Large deviations for processes with independent increments. *Ann. Probab.*, 15(2):610–627.

PUHALSKII, A. (1991). On functional principle of large deviations. In Sazonov, V. and Shervashidze, T., editors, *New Trends in Probability and Statistics*, volume 1, pages 198–218. VSP/Moks'las.

PUHALSKII, A. (1993). On the theory of large deviations. *Theory Probab. Appl.*, 38(3):490–497.

PUHALSKII, A. (1994a). Large deviations of semimartingales via convergence of the predictable characteristics. *Stochast. Stochast. Rep.*, 49:27–85.

Puhalskii, A. (1994b). The method of stochastic exponentials for large deviations. *Stochast. Proc. Appl.*, 54:45–70.

Puhalskii, A. (1995a). Large deviation analysis of the single server queue. *Queueing Systems*. (to appear).

Puhalskii, A. (1995b). Large deviations of semimartingales: a maxingale problem approach. I. Limits as solutions to a maxingale problem. (submitted).

Puhalskii, A. (1995c). Large deviations of the statistical empirical process. In *Frontiers in Pure and Applied Probability*. TVP. Proceed. of the Russian–Finnish symposium, 1993.

Radavichius, M. (1983). On the probability of large deviations of maximum likelihood estimators. *Sov. Math. Dokl.*, 27(1):127–131.

Radavichius, M. (1991). From asymptotic efficiency in minimax sense to Bahadur efficiency. In Sazonov, V. and Shervashidze, T., editors, *New Trends in Probab. and Statist.*, volume 1, pages 629–635. VSP/Mokslas.

Rao, C. (1963). Criteria of estimation in large samples. *Sankhya*, 25:189–206.

Rockafellar, R. (1970). *Convex Analysis*. Princeton University Press.

Rubin, H. and Rukhin, A. (1983). Convergence rates of large deviations probabilities for point estimators. *Statistics and Probability Letters*, 1:197–202.

Sanov, I. (1957). On the probability of large deviations of random variables. *Mat. Sbornik*, 42. (in Russian) ; English translation: Selected Translations in Mathematical Statistics and Probability, 1961, V. 1, pp. 213-244.

Schwartz, L. (1973). *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*. Oxford University Press.

Shiryaev, A. and Spokoiny, V. (1995). *Statistical Experiments and Decisions: Asymptotic Theory*. Springer. (in printing).

Sievers, G. (1978). Estimates of location: a large deviations comparison. *Ann. Statist.*, 6:610 – 618.

Strasser, H. (1985). *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. de Gruyter.

Varadhan, S. (1966). Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19(3):261–286.

Varadhan, S. (1984). *Large Deviations and Applications*. SIAM.

Vervaat, W. (1988). Narrow and vague convergence of set functions. *Statist. & Probab. Lett.*, 6(5):295–298.

Wieand, H. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.*, 4:1003–1011.